Clinical Gynecologic Endocrinology and Infertility

Contents

Author

Preface

Part I: Reproductive Physiology

Chapter 1: Molecular Biology for Clinicians
A basic guide to the language of molecular medicine.

Chapter 2: Hormone Biosynthesis, Metabolism, and Mechanism of Action
How hormones are formed and metabolized, and how hormones work.

Chapter 3: The Ovary — Embryology and Development
The ovary from conception to adult function; correlation of morphology with reproductive and steroidogenic function.

Chapter 4: The Uterus
Embryology, histology, and endocrinology of the uterus and menstruation. Anatomical abnormalities and leiomyomata.

Chapter 5: Neuroendocrinology
How reproductive events are perceived, integrated, and acted upon by the central nervous system.

Chapter 6: Regulation of the Menstrual Cycle
The cyclic changes of ovarian and pituitary hormones and growth factors, and what governs those changes.

Chapter 7: Sperm and Egg Transport, Fertilization, and Implantation
Mechanisms of the days just before and after conception.

Chapter 8: The Endocrinology of Pregnancy
The steroid and protein hormones of pregnancy.

Part II: Clinical Endocrinology

Chapter 9: Normal and Abnormal Sexual Development
Normal and abnormal sexual differentiation, and the differential diagnosis of ambiguous genitalia.

Chapter 10: Abnormal Puberty and Growth Problems
The physiology of puberty, and abnormalities that produce accelerated or retarded sexual maturation and growth problems in adolescents.

Chapter 11: Amenorrhea
Differential diagnosis of amenorrhea of all types utilizing procedures available to all clinicians. The problems of galactorrhea and pituitary adenomas, exercise and amenorrhea.

Chapter 12: Anovulation and the Polycystic Ovary
How loss of ovulation can occur and the clinical expressions of anovulation. The polycystic ovary and hyperinsulinemia.

Chapter 13: Hirsutism
The biology of hair growth; the evaluation and management of hirsutism.

Chapter 14: Menstrual Disorders
Medical problems linked to menstruation: the premenstrual syndrome, dysmenorrhea, menstrual headache, catamenial seizures, premenstrual asthma, and catamenial pneumothoraces.

Chapter 15: Dysfunctional Uterine Bleeding
A physiologic basis for medical management without primary surgical intervention.

Chapter 16: The Breast

Chapter 17: Menopause and the Perimenopausal Transition
Physiology of the menopause; long-term effects of hormone therapy on cognition, cardiovascular disease, and osteoporosis.

Chapter 18: Postmenopausal Hormone Therapy
A clinical guide for postmenopausal hormone therapy and patient management.

Chapter 19: Obesity
The physiology of adipose tissue, and the problem of obesity.

Chapter 20: Reproduction and the Thyroid
Normal and abnormal thyroid function, including a consideration of the thyroid gland in pregnancy.

Part III: Contraception

Chapter 21: Family Planning, Sterilization, and Abortion
The modern efficacy and use of contraception; the clinical methods and problems with sterilization and abortion.

Chapter 22: Oral Contraception

Chapter 23: Long-Acting Methods of Contraception
The advantages and disadvantages of specific long-acting contraceptive methods.

Chapter 24: The Intrauterine Device (IUD)
The characteristics of the IUD, an excellent method of contraception.

Chapter 25: Barrier Methods of Contraception
Important information for the use of the diaphragm, the cervical cap, the sponge, spermicides, and condoms.

Part IV: Infertility

Chapter 26: Female Fertility
An approach to the problem of infertility. The proper diagnostic tests and their correct interpretation.

Chapter 27: Recurrent Early Pregnancy Losses
The evaluation and management of recurrent spontaneous losses in early pregnancy.

Chapter 28: Ovulation
Diagnosis and suitable treatment for the individual patient.

Chapter 29: Male Infertility
Principles of male infertility, including analysis of semen, treatment, and therapeutic insemination.

Chapter 30: Induction of Ovulation
Programs, results, and complications for clomiphene, bromocriptine, gonadotropins, and GnRH administration; a reconsideration of the ovarian wedge resection.

Chapter 31: Assisted Reproduction
An overview of the assisted reproduction technologies.

Chapter 32: Ectopic Pregnancy
The diagnosis and treatment, both medical and surgical, of ectopic pregnancy.

Appendix
Selected laboratory values and the International System of Units (SI Units)
Preface

A quarter of a century later, the millennium and our 6th edition! The impact of our book has been one of the most gratifying experiences in our careers. We know it has helped many readers to successfully overcome examinations, but, most rewarding of all, is the recognition of the effect on patient care derived from the clinical usefulness of our text. An appreciation for this impact carries with it a profound sense of obligation. We have always understood the importance of “getting it right” and “causing no harm.” For that reason, every edition is the culmination of a tremendous effort, an effort increasingly difficult with the amazing growth of knowledge and communications. In 1972, it was a Royal portable typewriter. At the millennium, we are armed with a personal computer and a modem, Word 97 for the text, EndNote for references, and MEDLINE for searches.

The book grows progressively thicker, reflecting the explosion of knowledge in reproductive biology and medicine. Nevertheless, we believe we have stayed true to our objective: to provide both factual knowledge and systematic approaches to clinical problems. Our text continues to be an expression and formulation of our collective professional activities as teachers, investigators, and most importantly, as clinicians in the field of reproductive endocrinology and infertility. We have tried to retain the style and clinical relevance of our first edition, but at the same time to provide all the new information. This basic knowledge is the foundation for the physiologic principles that in turn are the foundation for our methods of clinical management.

As always, our book is dedicated to the improvement of patient care, and we hope that it will aid all readers in accomplishing that goal. We are grateful to those who have taught us, and most of all, we thank our patients, who are our best teachers.

Leon Speroff, M.D.
Portland, Oregon

Robert H. Glass, M.D.
San Francisco, California

Nathan G. Kase, M.D.
New York, New York

P.S. The cover color of the 6th edition is a nostalgic return to Yale blue. The symbol on the cover is the Macedonian Star, from the days of Philip of Macedon and Alexander the Great.
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.

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 Nobel prizes are not awarded posthumously.

DNA replication involves many enzyme systems. DNA polymerase was isolated in 1958, and 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.

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 for the new 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.

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 1 pair of sex chromosomes. All somatic cells are diploid—23 pairs of chromosomes. Only gametes are haploid with 22 autosome chromosomes and 1 sex chromosome. The chromosomes vary in size, and 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.

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 cross-overs 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. Cross-overs, 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 4 haploid cells.

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 4 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 3 major components: the deoxyribose sugar, a phosphate group, and a nucleic acid base. The phosphate-sugar linkages are asymmetric; the phosphorous 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.

How can a cell’s DNA, which stretched out measures nearly 2 meters 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 is, therefore, measured by the number of base pairs, e.g., a 4800-bp fragment (a 4.8 kb fragment). It is estimated that we have 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 reform 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 **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 3 billion base pairs in each haploid human genome; in the double stranded helix DNA, there are 6 billion nucleotides, and there are an estimated 60,000 to 150,000 genes, the smallest functional unit of inherited information. Genes account for only about 3% 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 3 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 20 Amino Acids in Proteins

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Phenylationine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
</tbody>
</table>

### The mRNA Genetic Code

<table>
<thead>
<tr>
<th>First Position (5¢ end)</th>
<th>Second Position (3¢ end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>Phe Ser Tyr Cys U</td>
</tr>
<tr>
<td></td>
<td>Phe Ser Tyr Cys C</td>
</tr>
<tr>
<td></td>
<td>Leu Ser Stop Stop A</td>
</tr>
<tr>
<td></td>
<td>Leu Ser Stop Trp G</td>
</tr>
<tr>
<td>C</td>
<td>Leu Pro His Arg U</td>
</tr>
<tr>
<td></td>
<td>Leu Pro His Arg C</td>
</tr>
<tr>
<td></td>
<td>Leu Pro Gln Arg A</td>
</tr>
<tr>
<td></td>
<td>Leu Pro Gln Arg G</td>
</tr>
<tr>
<td>A</td>
<td>Ile Thr Asn Ser U</td>
</tr>
<tr>
<td></td>
<td>Ile Thr Asn Ser C</td>
</tr>
<tr>
<td></td>
<td>Ile Thr Lys Arg A</td>
</tr>
<tr>
<td></td>
<td>Met Thr Lys Arg G</td>
</tr>
<tr>
<td>G</td>
<td>Val Ala Asp Gly U</td>
</tr>
<tr>
<td></td>
<td>Val Ala Asp Gly C</td>
</tr>
<tr>
<td></td>
<td>Val Ala Gln Gly A</td>
</tr>
<tr>
<td></td>
<td>Val Ala Gln Gly G</td>
</tr>
</tbody>
</table>

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.
Codon
A sequence of 3 bases in RNA or DNA that codes for a specific amino acid; the triplet codon.

With some exceptions, essentially one gene yields one protein. 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. 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 polyadenylate 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 promoter 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 mRNA, 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 3 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, 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 RNAases and at the other end, a polyadenylene 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). 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). Amino acids are brought into the process by specific transfer RNA molecules. The specific sequence of 3 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.

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 proopiomelanocortin to ACTH).

**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**

Any change in DNA sequence 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). 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 Edward syndrome, trisomy 21 Down syndrome, 47,XY 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.

**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.

**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. 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. 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 will be influenced by the gender of the parent providing the mutant gene or chromosome.

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) will cut 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. They 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.

**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, then digested by restriction enzymes to produce smaller fragments. The Southern blot method, named after E.M. Southern, determines the fragment sizes. The fragments are separated by electrophoresis. 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) 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 nitrocellulose membrane to which it binds. The salt solution keeps moving and is absorbed by the paper towels. The nitrocellulose 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 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.

Northern blotting refers to RNA processing. Northern because RNA is the opposite image of DNA. Extracted RNA is separated by electrophoresis and transferred to a cellulose 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.

Electrophoresis to separate 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. 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 nitrocellulose 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.

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.

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 (Thermus aquaticus) that is a thermophile (a hot water microbe) and was found in an out-of-the-way Yellowstone National Park hot spring called Mushroom Pool. This high-temperature polymerase allows automation of the process.

The technique of polymerase chain reaction has made possible the study of incredibly small amounts of DNA. Most 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).

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. 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 cosmids vectors, artificially produced combinations of phage and plasmid vectors. Very large fragments, up to 1000 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.

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).

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 being 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 frames. Once the sequence of a DNA fragment has been identified, the computer can utilize DNA and protein data bases 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, or simply polymorphisms, usually a benign variation. A polymorphism will be 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).

Minisatellites are a form of polymorphisms. 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. A polymorphism can serve as a genetic marker for a medically important gene.

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 is to sequence the 3 billion base pairs of the human genome, enough information to occupy 10,000 conventional floppy disks. Using the standard thin-layer vertical gel electrophoresis followed by autoradiographs of gel blots, it would require tens of millions of gels to accomplish this project. Shared data and computer analysis are necessary, and high speed automated sequencing technology must be utilized. Each laboratory will determine its small part of the sequence and then have to figure out where it fits in the big picture. When the computer finds an open reading frame, the probable amino acid sequence and the function can be predicted for the gene's protein. A genetic linkage map can serve as a foundation for locating disease sites and for integrating the genetic sequencing with biologic functions. Remember that there are anywhere from 60,000 to 150,000 genes. By the end of 1997, more than 60 million (of the total 3 billion) base pairs had been analyzed, and an electronic version is available on the World Wide Web at [http://www.ncbi.nlm.nih.gov/]. This represents about 50,000 genes. The complete genomic sequences had been determined for 141 viruses, 51 organelles, 2 eubacteria, and one eukaryote. A complete map of the human genome is predicted to be available by 2005. Soon, we will be able to have in our possession a personal CD that contains our individual complete genetic blueprint.

The U.S. Department of Energy maintains a Web site that provides basic information and useful links to other sites regarding the human genome project:

[http://www.ornl.gov/techresources/human_genome/publicat/publictions.html]

Genetic Maps. By 1997, about 6000 polymorphic markers had been identified to localize genes that predispose to disease. Using this method, a gene can be targeted for more specific localization.

Physical Maps. Genes of interest are localized by comparing overlapping DNA fragments. This isolates chromosomal segments for sequencing.

Sequence Maps. The challenge of the Human Genome Project can be appreciated by considering that technology is capable of reading about 36,000 DNA letters per day, and there are about 130 million DNA letters in the average human chromosome. Although the genetic and physical maps are nearing completion, complete
The chromosomal locations of genes responsible for hormone production are rapidly being 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.

But molecular progression 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 children avoided because of a bad match of genetic predispositions. Society will have to develop 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.

Clinical Applications

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 heterogeneous collection of mutations. Molecular diagnosis is further challenged by the need to not only find a subtle change in a gene but 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 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 17 has been isolated and cloned. Molecular technology was important in the characterization of the gene, which had been isolated and 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 anti-müllerian 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 Y chromosome carries a genetic locus responsible for differentiation of the testes, called the testes-determining factor. The theory that this is a male-specific histocompatibility antigen (the H-Y antigen) has been abandoned. Using material from intersex patients, maps of the Y chromosome have placed the testes-determining factor on the Y chromosome. This has been accomplished, and recombinant gonadotropins are now available. The gene for gonadotropin-releasing hormone on the short arm of chromosome 17 has been isolated and cloned. Molecular technology was important in the characterization of the gene, which had been isolated and 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 anti-müllerian 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. 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 oncoproteins and many different pathways of action, all of which result in a proliferative state. The mutations that activate these genes either lead 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.

There are also antioncogenes in normal cells, growth-suppressing genes that must be inactivated before tumors can grow. 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 antioncogene.

Cancer, therefore, is a genetic disease, but regulation of normal growth involves a complex system that takes a long time to overcome. 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 antibody for an oncogene, targeting the cancer cells.

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.

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 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, donate hepatocytes, or kill 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.

Chapter References


Online Resources

List of biochemical genetics tests by diseases, produced by the University of California, San Diego: http://biochemgen.ucsd.edu/wbgtests/dt-dst.htm

The U.S. National Library of Medicine search service for MEDLINE and other databases, including genomes, and DNA and protein sequences: http://www.ncbi.nlm.nih.gov/


Copyright ©


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. Words that are now encountered frequently are paracrine, autocrine, and intracrine, depicting 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.

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. M. Sutherland 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 new 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 purpose of this binding is not totally clear. The biologic activity of a hormone may be limited by binding in the blood, thereby avoiding extreme or sudden reactions. In addition, binding may prevent 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 will be metabolized without ever having the chance to produce an effect. Unlike estradiol, other hormones, such as testosterone, are metabolized and altered within the cell in which an effect is produced. In the latter case, a steroid is released into the bloodstream as an inactive compound. Clearance of steroids from the blood varies according to the structure of the molecules.

Thus, a steroid hormone has a varied career packed into a short lifetime. We are now ready to review the important segments of this lifespan in greater detail.

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 ring is benzene, two rings napthalene, and three rings phenanthrene; add a cyclopentane (5-carbon ring), and you have the perhydrocyclopentanephenanthrene structure of the steroid nucleus.

The sex steroids are divided into 3 main groups according to the number of carbon atoms they possess. The 21-carbon series includes the corticoids and 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 6 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 (a) (dotted line) and beta (b)(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 a position at C-17 rather than in the b position.

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 1, 2, or 3 double bonds: -ene, -diene, and -triene. Following the basic name, hydroxyl groups are indicated by the number of the carbon attachment, and 1, 2, or 3 hydroxyl groups are designated -ol, -diol, or -triol. Ketone groups are listed last with numbers of carbon attachments, and 1, 2, or 3 groups designated -one, -dione, or -trione. Special designations include: dehydro, elimination of 2 hydrogens; deoxy, elimination of oxygen; nor, elimination of carbon; delta or D, 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 5 major categories of lipoproteins according to their charge and density (rotation 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 (2/3) 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$_3$). HDL$_2$ is strongly associated with cardiovascular disease.

  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 degraded. 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
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 LDL-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, and 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).

Understanding the role of the cell surface receptors for the homeostasis of cholesterol (discussed later in this chapter), the work of the 1985 Nobel Laureates, Michael S. Brown and Joseph L. Goldstein, revolutionized our concepts of cholesterol, lipoprotein metabolism, and hormone action at the cell membrane. In their Nobel lecture, Brown and Goldstein paid tribute to cholesterol as the most highly decorated small molecule in biology.

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. Brown and Goldstein postulate that the LDL receptor evolved under dietary conditions of a lower fat intake, and that the high-fat, high-cholesterol modern diet suppresses the production of LDL receptors, thereby allowing cholesterol to rise to levels associated with cardiovascular disease.

**Steroidogenesis**

The overall steroid biosynthesis pathway shown in the figure is based primarily on the pioneering work of Kenneth. J. Ryan and his coworkers. 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 3 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 11b-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 hydroxy groups into ketones or ketones into hydroxy groups (dehydrogenase reactions).
3. Addition of OH group (hydroxylation reaction).
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.

Steroidalogenic enzymes are either dehydrogenases or members of the cytochrome P450 group of oxides. 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 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 the 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.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cellular Location</th>
<th>Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>P450scc</td>
<td>Mitochondria</td>
<td>Cholesterol side chain cleavage</td>
</tr>
<tr>
<td>P450c11</td>
<td>Mitochondria</td>
<td>11-hydroxylase, 18-hydroxylase, 19-methyloxidase</td>
</tr>
<tr>
<td>P450c17</td>
<td>Endoplasmic reticulum</td>
<td>17-hydroxylase, 17,20-lyase</td>
</tr>
<tr>
<td>P450c21</td>
<td>Endoplasmic reticulum</td>
<td>21-hydroxylase</td>
</tr>
<tr>
<td>P450aromEndoplasmic reticulum</td>
<td>Aromatase</td>
<td></td>
</tr>
</tbody>
</table>

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
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 this step. 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 steroidalogenic enzymes and accessory proteins. More acutely, cyclic AMP stimulates the hydrolysis of cholesteryl esters and the transport of free cholesterol to the mitochondria.

Most of the cholesterol used for steroid synthesis is derived from the mobilization and transport of intracellular stores. 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 substrates. 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 genes 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 polyepipctide (SAPI), 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 steroidalogen acute regulator (SIAR) protein. SIAR messenger RNA and protein are induced concomitantly with acute steroidogenesis in response to cyclic AMP stimulation; SIAR increases steroid production; and SIAR is imported and localized in the mitochondria. But most impressively, congenital lipoid adrenal hyperplasia (an autosomal recessive disorder) has been demonstrated to be a failure in adrenal and gonadal steroidogenesis due to a mutation in the SIAR gene, which results in premature stop codons. With this mutation, a low level of steroidogenesis is possible, even permitting feminization at puberty, but continuing trophic hormonal stimulation results in an accumulation of intracellular lipid deposits that destroy steroidogenic capability.

STIAR mediates the transport of cholesterol into mitochondria in adrenal and gonadal steroidogenesis, but not in the placenta and brain. It is synthesized in a precursor form as a 285-amino acid protein that has a 25-residue sequence cleaved from the NH terminus after transport into mitochondria. The mutant forms of SIAR undergo premature truncation that prevents this proteolytic cleavage. Mutations of the SIAR gene, located on chromosome 8p11.2, are the only inherited disorder of steroidogenesis not caused by a defect in one of the steroidalogenic enzymes. The absence of SIAR expression in placenta and brain indicates the presence of different mechanisms for cholesterol transport in those tissues.

It is important to note that once pregnenolone is formed further steroid synthesis in the ovary can proceed by one of 2 pathways, either via Δ 3-3b-hydroxysteroids or via the Δ 3-3-ketone pathway. The first (the Δ 3 pathway) proceeds by way of pregnenolone and dehydroepiandrosterone (DHA) and the second (the Δ 2 pathway) via progesterone and 17a-hydroxyprogesterone.

The conversion of pregnenolone to progesterone involves two steps: the 3b-hydroxysteroid dehydrogenase and Δ 3-3-ketosteroid 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 3b-hydroxysteroid dehydrogenase enzyme catalyzes both the dehydrogenation and isomerization reactions, and exists in 2 forms (type I and type II), encoded by two separate genes on chromosome 1 (the type II gene is expressed in the gonads and the adrenal glands). Once the Δ 3-ketone is formed, progesterone is hydroxylated at the 17 position to form 17a-hydroxyprogesterone. 17a-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 17a-ketone may be reduced to a 17b-hydroxyl to form testosterone by the 17b-hydroxysteroid dehydrogenase reaction. Both C-19 steroids (androstenedione and testosterone) are rapidly converted to corresponding C-18 phenolic steroids (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 5-3b-hydroxy C-19 steroid, DHA, by 17a-hydroxylation followed by cleavage of the side chain. With formation of the Δ 3-3-ketone, DHA is converted into androstenedione. It is thought that conversion of each of the Δ 3 compounds to their corresponding Δ 4 compounds can occur at any step; however, the principal pathways are via progesterone and DHA. Regardless of the precursor source, C-19 Δ 3-3-ketone substrates proceed to estrogens as noted above.

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 10. 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 local influencing factors.

Characterization of the P450c21 protein and gene cloning indicate that there is only one 21-hydroxylase enzyme, the P450c21 in the smooth endoplasmic reticulum. Two human P450c21 genes (the A and B genes) have been cloned (on chromosome 10), and the evidence indicates that only one (the B gene) is active. The molecular genetics of 21-hydroxylase deficiency indicate that the syndrome can be due to gene conversions of material in the active B gene to resemble material in the inactive A gene, as well as deletions in the P450c21 B gene. A conversion is similar to a cross-over in genetic effect. However, rather than appearing as a deletion or addition, the gene changes, but the number of gene copies does not change.

Aromatization is mediated by P450arom found in the endoplasmic reticulum. The human genome has one P450arom gene, located on chromosome 15q21.1, and designated as the CYP19 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. Aromatase transcription is regulated by several promoter sites that respond to cytokines, cyclic nucleotides, gonadotropins, glucocorticoids, and growth factors. Tissue-specific expression is regulated by tissue-specific promoters. Thus, this gene has alternative promoters that allow the expression of a variety of enzymes that vary in their expression in the ovary and nonreproduced expression in the placenta and adipose. Very specific inhibitors of P450arom have been developed which allow intense blockage of estrogen production, with clinical applications that include the treatment of breast cancer (e.g., anastrozole) and dysfunctional uterine bleeding. The aromatase complex also includes NADPH-cytochrome P450 reductase, a ubiquitous flavoprotein for transferring reducing equivalents.

The 17b-hydroxysteroid dehydrogenase and 5a-reductase reactions are due to non-P450 enzymes. The 17b-hydroxysteroid dehydrogenase is bound to the endoplasmic reticulum and the 5a-reductase to the nuclear membrane. The 17b-hydroxysteroid dehydrogenase enzymes convert estrone to estradiol, androstenedione to testosterone, and DHA to androstenediol, and vice versa. Four different isoforms have been identified, types I–IV. The type I enzyme is active in the placenta, converting estrone to estradiol. The type II and type IV enzymes form androstenedione and estrone from testosterone and estradiol, respectively, and the type III enzyme in the testis reduces androstenedione to testosterone.

The Two-Cell System

The two-cell system is a logical explanation of the events involved in ovarian follicular steroidogenesis. This explanation, first proposed by Falck in 1959, 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. FSII receptors are present on the granulosa cells.
2. FSII receptors are induced by FSII itself.
3. LH receptors are present on the theca cells and initially absent on the granulosa cells, but, as the follicle grows, FSII induces the appearance of LH receptors on the granulosa cells.
4. FSII 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.
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 Progesterone circulates in the following percentages: 2% unbound, 80% bound to albumin, 18% bound to transcortin, and less than 1% bound to SHBG. Binding in Transcortin, also called corticosteroid-binding globulin, is a plasma glycoprotein that binds cortisol, progesterone, deoxycorticosterone, corticosterone, and some of the other minor corticoid compounds.

SNBG 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. Genetic studies have revealed that the SHBG gene also encodes the androgen-binding protein present in the seminiferous tubules, synthesized by the Sertoli cells. 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. However, SHBG gene expression has now been identified in other tissues (brain, placenta, and endometrium), although a biologic significance has not been determined.

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 P450sc, P450c17, and 3b-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. 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 thecal cells. 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. Theca production of transforming growth factor can promote the growth of granulosa cells and FSH induction of LH receptors on the granulosa. 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 serve as substrate for aromatization to estrogen in the granulosa, whereas activin suppresses androgen synthesis. This important paracrine regulation of androgen production in the thecal cells by inhibin and activin is exerted primarily through modification of the expression of steroidogenic enzymes, especially P450c17.

After ovulation, the dominance of the luteinized granulosa layer is dependent on prevoluntary 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. 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.

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 in the liver. Another 10–30% 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.

<table>
<thead>
<tr>
<th>Free (Unbound)</th>
<th>Albumin-BoundSHBG-Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estrogen</td>
<td>1%</td>
</tr>
<tr>
<td>Testosterone</td>
<td>1%</td>
</tr>
<tr>
<td>DHA</td>
<td>4%</td>
</tr>
<tr>
<td>Androstenedione</td>
<td>7%</td>
</tr>
<tr>
<td>Dihydrotestosterone</td>
<td>1%</td>
</tr>
</tbody>
</table>

From Mendel

The circulating level of SHBG is inversely related to 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 (independent of age and weight). 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 II diabetes mellitus.

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. 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. Genetic studies have revealed that the SHBG gene also encodes the androgen-binding protein present in the seminiferous tubules, synthesized by the Sertoli cells. 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. However, 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: 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
SHBG 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. 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 many-fold greater than that of SHBG, the contribution of the albumin-bound fraction is 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 common precursors of estrogens. 17b-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. Estradiol 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. The work of Siiteri and MacDonald 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.

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.

\[
\text{MCR} = \text{Liters/Day} \\
\text{PR} = \text{MCR} \times \text{Concentration} \\
\text{PR} = \text{Liters/Day} \times \text{Amount/Liter} = \text{Amount/Day}
\]

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 microgram 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.

**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 use, except in home test kits that allow women to self-test for ovulation.

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 100 ng/dL. After ovulation, i.e., during the luteal phase, progesterone ranges from 500 to 2000 ng/dL. In congenital adrenal hyperplasia, progesterone blood levels can be as high as 50 times above normal.

Androgen Metabolism

The major androgen products of the ovary are dehydroepiandrosterone (DHA) 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 3 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 DHA and androstenedione comes from the adrenal gland; the other half of androstenedione is secreted by the ovary, but the other half of DHA 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 DHA) 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 intracelular androgen effects (specifically increased intracellular conversion of testosterone to dihydrotestosterone).

Reduction of the Δ1 unsaturation (an irreversible pathway) in testosterone is very significant, producing derivatives very different in their spatial configuration and activity. The 5α-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. 32 Thus, in women, the skin production of DHT may be predominantly influenced by androstenedione. DHT is by definition an intracrine hormone, formed and acting within target tissues. 32 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. 32
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. 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.

**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.

**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 2 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 in the nucleus 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).

**Other Receptors**
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 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 post transcriptional 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, has been demonstrated to act according to this general mechanism. Glucocorticoid, mineralocorticoid, and probably androgen receptors, when in the unbound state, reside in the cytoplasm and move into the nucleus after hormone-receptor binding. Estrogens and progestins 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 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 factors is a factor in the biologic response because the hormone response elements are abundant and, under normal conditions, are occupied only to a small extent. 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 that the long half-life of the estrogen hormone-receptor complex. Indeed, a major factor in the potency differences among the various estrogens (estradiol, estriol, 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. This process is called replenishment. Pregestosterone and clomiphene, on the other hand, limit tissue response to estrogen by blocking the replenishment mechanism, thus decreasing over time the concentration of estrogen receptors. Replenishment is very responsive to the available amount of steroid and receptors. Small amounts of receptor depletion and small amounts of steroid in the blood activate the mechanism.

Replenishment, 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 incredible nuclear traffic. The nuclear membrane contains 3000 to 4000 pores. A cell synthesizing DNA imports about one million histone molecules from the cytoplasm every 3 minutes. If the cell is growing rapidly, about 3 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; constant 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 conversion of high-affinity estrogen receptor sites to a rapidly dissociating form followed by loss of binding capacity, which is completed in about 6 hours. The rapid turnover of estrogen receptors has clinical significance. In pregnancy, where the concentration of estradiol is very great, it can be an important hormone, not just a metabolite.

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. In estrogen-regulated genes, the mechanism is different for androgen antiestrogen effects. Androgens do not involve depletion of estrogen receptors but in some way decrease estrogen-induced RNA activity in the cytoplasm.

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-dihydroxy vitamin D_3_, and retinoic acid; thus, these receptors are called a superfamily. 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 about 150 proteins, present in practically all species, from worms to insects to humans. Many are called orphan receptors because specific ligands for these proteins have not been identified.

The Estrogen Receptors

Two estrogen receptors have been identified, designated as estrogen receptor-alpha (ER-a) and estrogen receptor-beta (ER-b). The estrogen receptor-alpha was discovered about 1960, and the amino acid sequence was reported in 1986. The estrogen receptor-alpha is translated from a 6.8-kilobase mRNA that contains 8 exons derived from a gene on the long arm of chromosome 6. It has a molecular weight of approximately 66,000 with 595 amino acids. The receptor alpha half-life is approximately 4-7 hours, thus the estrogen receptor-alpha is a protein with a rapid turn over. The more recently discovered estrogen receptor-beta is encoded by a gene localized to chromosome 14, q22-q24, in close proximity to genes related to Alzheimer’s disease.

The estrogen receptors are divided into 6 regions in 5 domains, labeled A to F. The ER-b is 97% homologous in amino acid sequence with the alpha estrogen receptor in the DNA binding domain and 59% homologous in the hormone-binding domain. The full comparison is as follows:

<table>
<thead>
<tr>
<th>Region</th>
<th>ER-a Homology</th>
<th>ER-b Homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>The regulatory domain</td>
<td>17.5%</td>
<td>17%</td>
</tr>
<tr>
<td>The DNA-binding domain</td>
<td>97%</td>
<td>95%</td>
</tr>
<tr>
<td>The hinge</td>
<td>30%</td>
<td>26%</td>
</tr>
<tr>
<td>The hormone-binding domain</td>
<td>59.1%</td>
<td>59%</td>
</tr>
<tr>
<td>The F region</td>
<td>17.9%</td>
<td>17%</td>
</tr>
</tbody>
</table>

The following discussion represents information derived from studies of the ER-a. The hormone binding characteristics of the ER-a and the ER-b are similar, indicating that they respond in a comparable manner to the same hormones. There are differences, however; for example, phytoestrogens have a greater affinity for ER-b than for ER-a. Different genetic messages can result not only because of differences in binding affinity, but 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, ER-b may not be capable of activating gene transcription by means of TAF-1 (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-a, 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, and in ER-b, 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 9 cysteines in fixed positions, the two zinc fingers. This domain is essential for activation of transcription. Hormone binding induces a conformational change 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 finger tips 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-alpha is the hormone-binding domain (for both estrogens and antiestrogens), consisting of 251 amino acids (residues 302–553). In addition to hormone binding, this region is responsible for dimerization and contains 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. The hormone-binding domain of the steroid receptors contains a characteristic structure, containing helices that form a pocket (also referred to as a sandwich fold). After binding with a hormone, this pocket undergoes a conformational change that creates new surfaces with the potential to interact with co-activator and co-repressor proteins.

F Region

The F region of ER-a 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. 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.

Mechanism of Action

The steroid family receptors are predominantly in the nucleus even when not bound to a ligand, except for the androgen, 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 constantly diffuses out of the nucleus and is rapidly transported back in. When this shuttling is impaired, receptors are more rapidly degraded in the cytoplasm. Agents that inhibit dimerization (e.g., the pure estrogen antagonists) inhibit nuclear translocation and thus increase cytoplasmic degradation.
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 (also referred to as a sandwich fold). After binding with a hormone, this pocket undergoes a conformational change that creates new surfaces with the potential to interact with co-activator and corepressor proteins. Coactivators and corepressors — peptides that interact with the TAF areas of the receptor, also called adaptor proteins. Other transcription factors — peptides that interact with the polymerase enzyme and DNA. Chromatin factors — structural organizational changes that allow an architecture appropriate for transcription response.

The hormone-binding domain of the estrogen receptors contains a cavity surrounded by a wedge-shaped structure, and it is the fit into this cavity that is so influential in influencing the 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 at the same site 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. 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, 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. 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-alpha can form dimers with other alpha receptors (homodimers) or with an estrogen receptor-beta (heterodimer). Similarly, the estrogen receptor-beta 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 express 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. The zinc fingers on the various hormone receptors are not identical. These fingers of amino acids are thought to 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 (called 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 and androgens, one for estrogen, one for vitamin D3, and one for thyroid/retinoic acid. 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 3 different RNA polymerases (designated as 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. 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.

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.
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 (called adaptor proteins) that activate or suppress the TAF areas, either by acting on the receptors or on DNA and most of the genes regulated by estrogen 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.

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 estril. 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 conformation that allows TAFs to accomplish its 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.

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. 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.

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

**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. The steroid receptor superfamily members are phosphoproteins. Phosphorylation follows steroid binding and occurs in both the cytoplasm and nucleus. This phosphorylation is believed to enhance activity of the 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 TGF-a can activate the estrogen receptor in that activation is ligand independent. The response to growth factors can be blocked by pure antiestrogens (suggesting that a strong antagonist blocks the receptor 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-a, FGF), and the growth factors in an autocrine fashion further activate the estrogen receptor.

**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 by 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-a and ER-b).
3. Conformational shape of the ligand-receptor complex, with effects on two important activities: dimerization and modulation of adapter proteins.

**Different Roles for ER-a and ER-b**

Male and female mice have been developed that are homozygous for disruption of the alpha estrogen receptor gene, “estrogen receptor-alpha knockout mice.” Both sexes with this knockout are infertile. Spermatogenesis in the 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 b-subunit synthesis is increased, but FSH secretion is at normal levels, indicating different sites of action for estrogen and inhibit. 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-b and low levels of ER-a. However, nongenomic actions of estrogen are also possible and can explain some of the estrogenic responses in a knockout model.

Differential expression of the alpha and beta receptors is likely in various tissues (e.g., ER-b is the prevalent estrogen receptor in certain areas of the brain and the cardiovascular system) resulting in different and selective responses to specific estrogens. Human granulosa cells from the ovarian follicle contain only ER-b mRNA; the human breast expresses both ER-a and ER-b. Some parts of the rat brain contain only ER-b, others only ER-a, and some areas contain both receptors.

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-a and a given site of the estrogen response element, whereas estradiol inhibits gene transcription with ER-b in this same system. 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-a and ER-b 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-a, but because ER-b lacks a similar TAF-1, such agents can be pure antagonists in cells that respond only to ER-b. ER-a and ER-b affect the peptide context of a cell, especially coactivators and corepressors, differently.
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 phosphorylation). The progesterone receptor (in a fashion similar to the estrogen receptor) has two major forms, designated the A and B receptors. These two forms are a consequence of transcription from distinctly different promoters, in a complex system of transcription regulation. 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. 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 (164 amino acids) referred to as the B-upstream segment (BUS).

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. In the absence of hormone binding, the C-terminal region of the progesterone receptor exerts an inhibitory effect on transcription. 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.

Proggestational 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.

A and B are expressed in approximately equal 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. 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 through 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.

The broad activity of A in regards to all steroids suggests that A regulates 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, A either 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.

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. 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. It is likely that the A and B forms of the androgen receptor have functional differences; however, this remains to be characterized. 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. Androgens and progestins can cross react for their receptors but do so only when present in pharmacologic concentrations. Progestins not only compete for androgen receptors but also compete for the metabolic utilization of the 5a-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 testicular feminization (androgen insensitivity) represents a congenital abnormality in the androgen intracellular receptor (about 200 unique mutations have been identified). 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. Thus, testicular feminization is an X-linked disorder. Molecular studies of patients with testicular feminization have indicated a deletion of amino acids from the steroid-binding domain due to nucleotide alterations in the gene that encodes the androgen receptor. 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 testicular feminization. In the middle is a spectrum of clinical presentations representing varying degrees of abnormal receptors and binding. While at the other end, it has been suggested that about 25% of infertile men with normal genitalia and normal family histories have azoospermia due to a receptor disorder. 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.

Nongenomic Actions of Steroid Hormones

The genomic effects of steroid hormones are characterized by a relatively slow response time of 1 hour or longer. However, some steroid hormone effects are immediate, within a few seconds, and nongenomic mechanisms must be operative in order to achieve such rapid responses. These rapid responses are also unaffected by inhibitors of gene transcription or protein synthesis. Rapid actions have been reported for all steroid hormones and include calcium and sodium transport across membranes, neural effects, and certain oocyte and sperm reactions. The messenger and effector systems utilized vary from cell to cell and from steroid to steroid. Specific cell membrane receptors have been identified for various steroids; however, it has been difficult to demonstrate physiologic roles for these binding sites. Nevertheless, investigation thus far indicates that steroid hormones can bind to membrane receptors and trigger rapid changes in electrolyte transport systems. Estrogen-induced vasodilatation in the coronary arteries is believed to be mediated, at least in part, through a nongenomic calcium flux mechanism.

Agonists and Antagonists

An agonist is a substance that stimulates a response. An antagonist completely 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, RU486, 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 the normal receptor-DNA processing and a failure to replenish hormone receptors, resulting in eventual depletion.

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 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. In addition, progestins induce enzyme activity that converts the potent estradiol to the impotent estrone sulfate, which is then secreted from the cell. Androgens do block the actions of estrogen, but the mechanism is not entirely clear. Rather than a direct impact on estrogen receptor levels, the action is directed to gene activity subsequent to estrogen-receptor binding. High levels of androgen can produce estrogen and progestational effects by binding to the estrogen and progesterone receptors.
Antiestrogens

Currently, there are two groups of antihormones: pure antihormones 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 benzothiazepines such as raloxifene). The pure antihormones have a bulky side chain that, with only a little imagination, can be pictured as an obstruction to appropriate conformational changes. An ideal antihormone 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.

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 100–1000 times greater than that of tamoxifen. Thus, tamoxifen must be present in a concentration 100–1000 times greater than estrogen to maintain inhibition of breast cancer cells. In vitro studies demonstrate that this action is not cytotoxic, 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, antihormonoscopic message predominates is determined by what promoter elements are present in specific cell types.

There have been many clinical trials with adjuvant treatment of breast cancer with tamoxifen, and many are still on going. Overall, the impact of tamoxifen treatment on breast cancer can be summarized as follows: disease-free survival is prolonged. There is an increased survival at 5 years of approximately 20%, 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 HDL-cholesterol and 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. 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 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 currently used. Tamoxifen causes a decrease in antithrombin III, and there has been a small increase in the incidence of thrombo-embolism observed in tamoxifen-treated patients compared with controls. Also, however, in the overall case of randomized trials, no significant cardiac or vascular increase in mortality was noted in tamoxifen-treated women.

All too often, the agonistic, antihormonoscopic action of tamoxifen is featured, and the estrogenic, agonistic action is ignored. In the 1980s, it was reported that human endometrial cancer transplanted into mice would grow more rapidly during tamoxifen therapy, although the growth of breast cancer cells would be inhibited. This growth in response to tamoxifen can be duplicated in laboratory culture preparations of endometrial cancer cell lines.

There have now been many reports of endometrial hyperplasia, endometrial polyps, and endometrial cancer occurring in women receiving tamoxifen treatment. 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 antihormonoscopic activity. How can tamoxifen be both an estrogen agonist and an estrogen antagonist?

Tamoxifen Mechanism of Action

Tamoxifen and TAF-1 areas can both activate transcription, but TAF-2 activites 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 ability is due to competitive inhibition of the estrogen-dependent activation of TAF-2.

An estrogen-associated protein 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.

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.

The contact sites of estrogens and antihormones with the estrogen receptor are not identical. When an antihormone binds to the estrogen receptor, the conformational changes that are induced after the ability of the estrogen receptor-antihormone complex to modulate transcriptional activity. The relative agonist—antagonist activity is determined by the specific conformation achieved by the specific antihormone.

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 antihormone 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. 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 now 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. Antihormones have no effects on TAF-1 dependent transcription in breast cells.
The nature of the estrogen receptor.

The cell context of protein coactivators and corepressors.

The properties of the ligand.

Tamoxifen and clomiphene inhibit protein kinase C activity (phosphorylation).

Modulation by growth factors and agents that affect protein kinases and phosphorylation.

Tamoxifen and estrogen have opposing effects on growth factors:

Tamoxifen inhibits calmodulin-dependent cyclic AMP phosphodiesterase, by binding to calmodulin.

Changes in pharmacology and metabolism of tamoxifen might occur so that cells acquire the ability to metabolize the antagonist to greater agonist activity. Some of these actions (especially inhibition of protein kinase activity and stimulation of TGF-b production) occur independently of tamoxifen binding to the estrogen receptor, and thus, estrogen receptor-negative tumors can be affected by these actions.

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. 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.

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. 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.

2. Variant and mutant estrogen receptors

Mutations in resistant breast tumors are infrequent and are unlikely to account for resistance. 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. This mechanism operates through the phosphorylation of the estrogen receptor or proteins involved in estrogen receptor-mediated transcription.

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.

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 been reported in only one study of patients with resistant tumors.
The Cyclic AMP Mechanism

Receptors from this membrane family are also found in the membranes of lysosomes, endoplasmic reticulum, Golgi complex, and in nuclei. The regulation of these receptor proteins is coupled to an active agent, an intracellular messenger. The major intracellular messenger molecules are cyclic AMP, inositol 1,4,5-triphosphate (IP3), and diacylglycerol (1,2-DG), calcium ion, and 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.

Another possible mechanism for pure antiestrogens involves a binding protein for the insulin-like growth factors. In a breast cancer cell line, ICI 182,780 inhibited growth and increased transcription of the IGFBP-3 gene. Estradiol did the opposite. In the uterus, tamoxifen and estrogen suppress IGFBP-3 production, while the ICI antagonist markedly increases IGFBP and causes uterine involution.

Because these agents function in a different manner than tamoxifen, it is not surprising that tamoxifen-resistant tumors respond to these agents.

Selective Estrogen Agonists/Antagonists (Selective Estrogen Receptor Modulators)

Agents such as raloxifene and droloxifene have antiestrogenic activity in the uterus as well as in the breast, and at the same time exert agonistic effects in certain target tissues. Raloxifene inhibits bone resorption and improves lipid levels (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 different target systems (such as bone) and to avoid unwanted actions (such as endometrial stimulation).

The Antiprogestin RU486

Both progesterone and the antiprogestins, RU486 (mifepristone) and ZK98299 (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. RU486 has some agonistic activity due to its ability to activate certain, but not all, of the transcription activation functions on the progesterone receptor. 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 RU486, the first successful antiprogestin was produced by scientists at Roussel Uclaf, a pharmaceutical company in Paris. RU486 is a 19-nortestosterone derivative. The dimethyl (dimethylaminophenyl) side chain at carbon 11 is the principal factor in its antiprogesteral action. There are three major characteristics of its action which are important: a long half-life, high affinity for the progesterone receptor, and active metabolites.

The affinity of RU486 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 androgen because of its low-affinity binding to the androgen receptor.

RU486 also binds to the glucocorticoid receptor, but higher doses are required to produce effects. The binding affinity of RU486 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, 1000-fold higher than progesterone. This allows titration of clinical effects by adjustments of dose.

Both progesterone and RU486 induce conformational changes with the progesterone receptor, especially in the hormone-binding domain. 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 RU486).

RU486 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 androgen antagonist, 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 released 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 either 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 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 2 regulatory subunits and 2 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. 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.

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.

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. This is 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 (cyclic guanosine 3'5'-monophosphate) may 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 may 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. Studies of this receptor family indicate that prolactin operates through various signal transduction mechanisms, including ion channels and nuclear kinase activation.

Gonadotropin releasing hormone (GrRH) is calcium dependent in its mechanism of action and utilizes IP3 and 1,2-DG as second messengers to stimulate protein kinase activity. 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. 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 polyphosphatidylinositol, specific phospholipids in the cell membrane. Activation of this enzyme by hormone binding to its receptor leads to the generation of 2 intracellular messengers, inositol triphosphate (IP$_3$) and diacyl glycerol (DAG), which initiate the function of the 2 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 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 4 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 3-dimensional conformation that provides ligand specificity. The transmembrane domains are not highly conserved (thus differing in make up). 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 4 major components.

1. Autocrine and paracrine regulation factors.
2. Heterogeneity of the hormone.
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 the 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 epithelial 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. Activin and inhibin are disulfide-linked dimers composed of peptide subunits (one alpha subunit and two beta subunits) as follows: 123

The 3 Forms of Activin:
- Activin A: Beta-A
- Activin AB: Beta-A Beta-B
- Activin B: Beta-B Beta-B

The 2 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 antimullerian 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 polypeptide, structurally unrelated to activin and inhibin, 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 FGF. Additional roles for FGF, secreted by the granulosa, include modulation of enzyme activity involved in the physical development of the corpus luteum.

The Insulin-Like Growth Factors

The insulin-like growth factors (also called somatomedins) are single-chain polypeptides that resemble insulin in structure and function. 124 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, while 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 factors (FSH, LH, TSH, and HCG) are IGF-II in both the granulosa and the theca cells. Studies indicate that the activity of IGF-I with human androgenic 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 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. 125 IGF-I and IGF-II circulate in the blood in a concentration 1000 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 upon 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.

Orphan Receptors Involved in Steroidogenesis

Steroidogenic factor-1 (SF-1) and DAX-1 (a name that represents: Dosage-sensitive sex reversal-Adrenal 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. 126 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. 127 SF-1 also regulates genes that encode the gonadotropin subunits, as well as the GnRH receptor. 128 Thus, SF-1 is 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

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. 129 The various forms (isomers) arise in various ways, including different DNA promoter actions, alterations in RNA splicing, point mutations, and post-translational carbohydrate changes. The impact of the variations is to alter structure and metabolic clearance, thus affecting binding and activity. The isomers have different molecular weights, circulating half-lives, and biologic activities. Throughout the menstrual cycle, the amazing number of at least 20–30 isomers of both FSH and LH are present in the bloodstream. 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 heterodimers bind to specific receptors, while 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 up of the tropic hormones is an essential element in determining the ability of the hormone to mate with its receptor. The glycosylated subunits of the human gonadotropins (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. 133 All of the glycopeptides of the human species (FSH, LH, TSH, and HCG) share a common a-chain, an identical structure containing 92 amino acids. The b-chains (or the
b-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 glycoprotein hormone is determined by the b-subunit; hypogonadism has been reported due to single amino acid substitution in the LH b-subunit. 144

b-HCG is the largest b-subunit, containing a larger carbohydrate moiety and 145 amino acid residues, including a unique carbohydrate-terminal tail piece 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 carboxy-terminal region of b-HCG contains 4 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 b-subunit gene compared with the site in the LH b-subunit gene. The HCG b-subunit 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 b-subunits, because excess a-units can be found in blood and in tissue. Furthermore, the three dimensional structure of the b-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 a-subunit. 145 This conformational change is not completed until the subunits are fully united to produce the final whole hormone.

The half-life of a-HCG is 6–8 minutes, 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 a-subunit of 92 amino acids and a b-subunit of 118 amino acids. It has four carbohydrate side chains, two on each subunit. The b-subunit of LH consists of 121 amino acids. LH has 3 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. 146 Using recombinant DNA technology, it has been demonstrated that there is a single human gene for the expression of the a-subunit. The gene for the a-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 a-gene in both placenta and pituitary. The a-subunit gene is expressed in several different cell types, but the b-subunit genes are restricted in cell type. The TSH b-gene is expressed only in thyrotrophs regulated by thyroid hormone; the FSH b-gene is expressed in gonadotrophs regulated by GnRH, activin, inhibin and gonadal steroids; the LH b-gene, also expressed in gonadotrophs, is regulated by GnRH and unaffected by activin and inhibit. 147

The a-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 a-gene expression. In gonadotrophs, the GnRH signaling pathway for a-gene transcription utilizes phosphorylase stimulation of diacyl glycerol (DAG) and inositol phosphate (IP) that leads to a release of intracellular calcium stores. GnRH also stimulates the influx of calcium at the cell membrane. DAG, IP, and calcium work together to stimulate protein kinase C activity. Protein kinase regulation of the a 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 b-subunit is on chromosome 11p13, and in the pituitary, it is markedly influenced by activin. 148 Although FSH and LH both require GnRH stimulation, the FSH b-gene is not in that pituitary. In LH, LH is markedly influenced by activin, and the LH b-gene is influenced by GnRH and activin. Activin is further antagonized by inhibit, the first of these factors recognized to suppress FSH secretion. 149

The genes that encode for the b-subunits of LH, HCG, and TSH are located in a cluster on chromosome 19q13.3. There are 6 genes for the b-subunit of HCG, and only one for LH. 150 Transcription for the 6 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 b-HCG evolved relatively recently from the b-LH gene and the b-LH gene are 96% identical. 151 Only primates and horses have been demonstrated to have genes for the b-subunit of choriocarcinoma. In contrast to human choriocarcinoma, equine choriocarcinoma exerts both LH and FSH activities in many mammalian species because it contains peptide sequences in its b-subunit that are homologous to those in the pituitary gonadotrophs of other species. The equine b-chorionic gonadotropin gene is identical to the equine b-LH gene, and although the b-subunit genes have diverged by difficulties in maintaining glycoprotein-producing cell lines. The availability of choriocarcinoma cell lines, however, has allowed greater investigation of the b-HCG genes.

Although the b-subunit specifies the biologic activity of an individual glycoprotein, the combination of the a- and b-subunits is necessary for full hormonal expression. Furthermore, the a-subunit also plays an important role in accomplishing normal receptor binding and activation. 152 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 a-b-Pair.
Variations in Carbohydrate

The glycopeptide hormones can be found in the pituitary existing in a variety of forms, differing in their carbohydrate (oligosaccharides) make up. The isoform mixture of gonadotropins is influenced both quantitatively and qualitatively by GnRH and the feedback of the steroid hormones, producing post translational carbohydrate modifications.  

Certain clinical conditions may be associated with alterations in the usual chemical structure of the glycopeptides, resulting in an interference with the ability to bind to receptors and stimulate biological activity. In addition to deglycosylation and the formation of antigonadotropins, 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 biological activity. 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). 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 sialylated, shorter-lived isoforms. These changes are effects of both GnRH and estrogen.

The carboxyl-terminal 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. 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.

The circulating half-life of a gonadotropin is mainly proportional to the amount of sialic acid present. The higher content of sialic acid in FSH compared with LH accounts for the more rapid clearance of LH from the circulation (FSH half-life is several hours; LH half-life is about 20 minutes). HCG is highly sialylated, and accordingly, has a half-life of many 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. 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.

Prolactin is encoded by a single gene on chromosome 6, producing a molecule that in its major form is maintained in 3 loops by disulfide bonds. 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 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.

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 moieties can produce differences in biologic activity and immunoreactivity. However, the glycosylated form of prolactin is the predominant form of prolactin secreted into the circulation. 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 upon the site of action (e.g., decidua and placenta). 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 immunoactivity (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

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 which 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 biological 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; however, further effects on target tissue response also occur at sites distal to receptors.

It is believed that receptors are randomly inserted into the cell membrane after intracellular synthesis. The receptor may be viewed as having 3 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 1500 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. 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.

A similar process, called pinocytosis, utilizes cholesterol-rich membrane invaginations called caveoleae (far fewer in number and smaller in structure than the clathrin coated pits) for the internalization of small molecules and ions. This is another method of intracellular signaling in response to hormones, and many proteins involved in cell signaling have been detected in caveoleae: e.g., G proteins, kinases, and growth factor receptors. Caveolin is the major protein structural component of caveoleae. Nitric oxide, the important mediator of vascular events, resides in caveoleae and is regulated by tyrosine phosphorylation and interaction with caveolin.

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.

Cell membrane receptors can be randomly distributed in the cell membrane and transmit information to modify cell behavior. 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.

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. The low-density lipoprotein particle is a sphere. It contains in its center about 1500 molecules of cholesterol which 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 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. 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.
Different cell surface receptors and proteins contain similar structural parts. 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 pre-existing 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 5 recognizable domains.

1. NH₂-terminal of 292 amino acids, composed of a sequence of 40 amino acids repeated with some variation 7 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 may 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 steriodogenesis, reesterification for storage, use in membrane structures, or excretion. Excretion (release of free cholesterol into the circulation by means of the HDL mechanism) involves the cell surface caveolae. 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. The LDL receptor makes one round trip every 10 minutes during its 20-hour lifespan for a total of several hundred trips. Genetic defects in receptors for LDL lead to a failure in internalization and hyperlipidemia.

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 3 protein units: the receptor, a guanyl nucleotide regulatory unit, and a catalytic unit. The regulatory unit is a coupling protein, regulated by guanine nucleotides (specifically GTP), and therefore it is called GTP binding protein or G protein for short. The catalytic unit is the enzyme itself which converts 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) retuning 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 GDP.

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 proteins, 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. 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. The β- and γ-subunits are not all alike, and they exhibit selectivity for specific receptors. 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 4 subfamilies: Gs, Gi, Gq, and G12. Gs and Gq proteins mediate stimulatory events such as hormone secretion, whereas Gi proteins exert inhibition. The role of the G12 group is not yet certain. These multiple subunits allow great variability in function to be expressed by many different combinations.

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.

The G protein Receptors

The more than 200 receptors linked to G proteins are derived from a supergene family, presumably originating from a common ancestral gene. 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. Receptors that utilize the G proteins are inserted in membranes and consist of a long polypeptide chain that folds into 7 helices, the amino acid loops that connect the helixes 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 a gene on chromosome 2. The LH/HCG receptor is highly conserved in mammals; the human receptor is very similar to that of rat and bovine receptors. 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. Actually, two LH/HCG receptor genes have been isolated (Gene I and Gene II), Gene I from a lymphocyte library and Gene II from a human placental library. Gene II appears to be the major form expressed in the ovary.

The receptor for FSH is very similar to the LH/HCG receptor, but it is structurally distinct. Appropriately (for specificity), the extracellular segment contains the major sequence divergence. The FSH receptor gene is located on chromosome 2p21, 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

Mutations that alter the structure and activity of G proteins can result in disease. 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 Gα mutation. 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 Gα gene. Gα 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.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activating LH receptor</td>
<td>Precocious puberty in boys</td>
</tr>
<tr>
<td>Inactivating LH receptor</td>
<td>Male pseudohermaphroditism</td>
</tr>
<tr>
<td>Inactivating FSH receptor</td>
<td>Premature ovarian failure</td>
</tr>
<tr>
<td>Gα (stimulatory)</td>
<td>McCune-Albright syndrome</td>
</tr>
<tr>
<td>Gα (inhibitory)</td>
<td>Hypothyroidism</td>
</tr>
</tbody>
</table>
Coupling and Uncoupling, Desensitization

Another way to explain stimulating and inhibiting actions at the adenylate cyclase level focuses on the mechanism of coupling. LH stimulates adenylate cyclase in the corpus luteum and works through the coupling of stimulatory regulatory units to the catalytic units of adenylate cyclase. Progesterol binding to the aldosterone receptor involves the following 3 mechanisms:

1. Desensitization by autophosphorylation of the cytoplasmic segment of the receptor.
2. Loss of receptors by internalization, a slow relatively mechanism.
3. Uncoupling of the regulatory and catalytic subunits of the adenylate cyclase enzyme.

Chapter References


60. Rowan MT, Dihydrotestosterone is a peripheral paracrine hormone, Mol Endocrinol 5:42, 1991.


The great names of early Western medicine were Hippocrates, Soranus, and Galen. Although Aristotle (384–322 BC) referred to castration as a common agricultural practice, it was Soranus who provided the first anatomical 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 AD 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.

Galen was born in 130 AD in Pergamum, a Greek city in eastern Turkey, studied in Alexandria and became a famous practitioner and teacher of medicine 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. 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 1000 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 anatomical knowledge can be traced to the University of Padua, the famed Italian university where a succession of anatomists made important contributions. 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 Humanis 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. Fallopian (1534–1562), 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'Acquapendente, 1533–1619), a pupil of Fallopius, succeeded Fallopius as chair of anatomy at Padua and made major contributions to embryology. Studying the organ in the bird that contained eggs, Fabricius called it the "vary." 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 Regnir de Graaf published his great work on the female reproductive organs. De Mulierum Organis Generativi 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 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 spermatozoan 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 1853–54, 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 cycle 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 androgenesis 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, the germinal epithelium. 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. These are 1) the indifferent gonad stage, 2) the stage of differentiation, 3) the period of oogonial 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 mediulary mesenchymal tissue. Just below this ridge lies the mesonephric duct. This indifferent stage lasts about 7–10
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 3 different tissues: coelomic epithelium, mesenchyme, and mesonephric tissue. In one model, the gonad is 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 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, the somatic cells of the gonad are believed to arise from the mesonephros and not the coelomic epithelium. Ultrastructural studies have even suggested that both coelomic epithelial and underlying mesonephric cells provide the somatic cells that are destined to become follicular cells.

Some factors that are involved in the formation of the gonads include:

1. The presence of gonadotropic hormones, which stimulate the development of the gonads.
2. The presence of steroidogenic enzymes in the gonads.
3. The presence of specific receptors for the gonadotropic hormones.
4. The presence of specific genes that are involved in the differentiation of the gonads.

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. 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.

Testosterone synthesis begins at the 8th week of gestation, reaching 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 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. The Leydig cells differentiate (beginning week 8) from mesenchymal cells of the interstitial component surrounding the testicular cords. Thus, secretion of AMH (anti-müllerian hormone and testosterone) of the fetal testes, while the female phenotype is the result of an absence of these fetal gonadal products. AMH expression is under the control of the SRY gene, which is expressed in the embryonic testis. The SRY gene is located on the Y chromosome and is responsible for the development of the testis. If the SRY gene is present, the gonad will develop into a testis; if it is absent, the gonad will develop into an ovary.

The Leydig cells begin their proliferation during their migration. The germ cells are the direct precursors of sperm and ova, and by the 6th gestational week, on completion of the indifferent stage, these primordial germ cells have multiplied by mitosis to a total of 10,000. The 6th 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. 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 gene located on the Y chromosome. Currently, the best candidate for the testicular determining factor gene is located within a region named SRY, the sex-determining region on the Y chromosome. The protein product of the SRY gene contains a DNA-binding domain that activates gene transcription. Normal testis development requires not only the presence of the SRY gene but its interaction with other genes. Genes similar to SRY have been named SOX genes (the similarity is with the SRY box region that contains the DNA-binding sequence). 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. The traditional view assigns active gene control and expression for testicular differentiation and a more passive, “default” mode of development for the ovary. However, any process of differentiation requires gene expression, and therefore ovarian development, too, must involve genes and gene products.

The Leydig cells differentiate from mesenchymal cells of the mesonephric tissue that contain the germ cells and supporting cells derived from the coelomic epithelium and the mesenchyme of the gonadal ridge. The Leydig 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.
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 regression of fetal Leydig cells and the appearance of morphologically distinct adult cells during the peripubertal period raise many questions: what is the relationship between these two types, what is the mechanism for the regression and for the new appearance, and what is the origin of the adult cells?

The fetal spermatagonia, 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, probably explaining why male internal genitalia in true hermaphrodites are only on the side of the tests. 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. In the female, the loss of the wolffian system is due to the lack of locally produced testosterone.

The Stage of Follicle Formation

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. 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 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. 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 steroids is present in follicular fluid, presumably secreted by the cumulus cells, that activates oocyte meiosis.

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. 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. 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 by birth, and it appears as a fibrous streak.

The Stage of Oogonal Multiplication and Oocyte 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. 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 are either mesenchymal or epithelial in origin. These cells surround the oocytes that 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. The granulosa cells differentiate from coelomic epithelial or mesenchymal precursors (their specific origin is still disputed). 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 the pregranulosa layer to a cuboidal layer of granulosa cells. Further differentiation into a preantral follicle is expressed as more complete granulosa proliferation. Call–Exner body formation (coalescence to form an antrum) and occasionally a minor thecal layer system that differentiates from surrounding mesenchymal cells can be seen. Preantral follicles can be found in the 6th 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.

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
The ovary at birth and in the first year of life can contain cystic follicles of varying size, undoubtably stimulated by the reactive gonadotropin surge accompanying the withdrawal of the neonatal hypothalamus and pituitary from the negative feedback of fetoplacental steroids. 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. Pluripotent levels of follicle-stimulating hormone (FSH) peak at 20–23 weeks, and circulating levels peak at 28 weeks. 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 occur, apparently not dependent upon gonadotropins. 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. Hypophysectomy of a fetal monkey is followed by an increase in oocyte loss by atresia.

The Neonatal Ovary

The total cortical content of germ cells falls to 1–2 million by birth as a result of prenatal oocyte depletion. 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.

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. 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 probably 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 one year of age, but sometimes later. 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.

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. Ovarian cysts can also be occasionally detected in fetuses by ultrasonography.

Interference with the postnatal rise in gonadotropins in monkeys is associated with disturbances in normal hypothalamic-pituitary function at puberty. 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. 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 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. These small unilocular ovarian cysts are not clinically significant. The process of atresia with an increasing contribution of follicular remnants to the stromal compartment yields progressive ovarian enlargement during childhood, about a ten-fold increase in weight. 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 mRNAs and protein. Furthermore, ovariectomy in prepubertal monkeys indicates that the pubertal suppression of GnRH and gonadotropins is partially dependent on the presence of oocytes, suggesting some functional activity of the ovary in childhood.

The Adult Ovary

At the onset of puberty, the germ cell mass has been reduced to 300,000 to 500,000 units. 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 remain. In the last 10–15 years before menopause, there is an acceleration of follicular loss. This loss correlates with a subtle but real increase in FSH and decrease in inhibin B as well as insulin-like growth factor-I (IGF-I). The accelerated loss is probably secondary to the increase in FSH stimulation. Fewer follicles grow per cycle as a woman ages, and cycles are at first shorter due to more rapid follicular growth prior to ovulation (because of the elevated FSH levels) and then longer as anovulation becomes more common. These changes, including the increase in FSH (which is probably due to the decrease in inhibin B), all reflect the reduced quality and capability of aging follicles. However, the rise in FSH due to a decrease in inhibin B may be the consequence of a decreasing number of follicles in each cohort of active follicles.

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. Indeed, the process is a consequence of an orderly expression of key gene products that either promote or repress the apoptotic events.

Unlike nonprimate species, human ovaries and nonhuman primate ovaries are innervated by sympathetic and sensory neurons. This neuronal network innervates the ovarian vasculature, interstitial tissue, and developing follicles. These neuronal cells produce catecholamines and nerve growth factor. The precise function for this unique primate ovarian nervous system is not known. However, nerve fibers (not in an organized network) are present in nonprimate ovaries. Vasoactive intestinal peptide derived from these nerve fibers suppresses follicular atresia (apoptosis) in a mechanism that also involves IGF-I.

During the reproductive years, the typical cycle of follicle maturation, including ovulation and corpus luteum formation, will be 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 Chapter 5 and Chapter 6. For the moment, our attention will be 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 1000 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.

It is now believed that the time that elapses in progressing from a primary follicle to ovulation is approximately 85 days. 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.

The antrum first appears as a coalescence of numerous intragranulosa cavities called Call–Exner bodies, that were described by Emma Call and Siegmund Exner in Vienna, in 1875. Emma Call was one of the first women physicians in the United States. 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 thecal 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.

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 thecal 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 distension 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 dissection of the antrum with rupture and extrusion of the oocyte in its cumulus, are multiple and complex (discussed in Chapter 6). Repeated evaluation of intraluminal 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. Failing a new enlarging source of LH-like 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 support 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. During the adult ovarian cycle, there is an infiltration of...
Anatomical knowledge of the uterus was slow to accumulate. Papyrus writings from 2500 BC 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 ovaries, 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 (4th century BC), the great anatomist in Alexandria and the originator of scholarly dissection, recorded the different positions of the uterus. Soranus of Ephesus (98–138 AD) 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 leiomyomata, and treated prolapse with pessaries.

Herophilus and Soranus were uncertain about the function of the fallopian tubes, but Galen, Rufus, and Aetius guessed correctly 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. The writings of Galen (130–200 AD) represented the knowledge of medicine for over 1000 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 anatomical study of the female genitalia. Berengarius (Giacomo Berengario da Carpi) was the first anatomist to work with an artist. His anatomical 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, one 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 3rd fetal month, except for nonfunctional vestiges.

Hormonal control of mammalian somatic sex differentiation was established by the classic experiments of Alfred Jost. 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 perhaps even the brain. The critical factors in determining which of the duct structures stabilize or regress are the secretions from the testes: AMH (anti-Müllerian hormone, also known as müllerian inhibiting substance or müllerian inhibiting factor) and testosterone.

AMH is a member of the transforming growth factor-b 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, 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. 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. The fallopian tubes, uterus, and the upper portion of the vagina are created by the fusion of the müllerian ducts by the 10th 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 struma 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 Bartlemeez and Markee in the subhuman primate. From these data a description of menstrual physiology has developed based upon specific anatomic and functional changes within glandular, vascular, and stromal components of the endometrium. 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 and, 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.

### 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 the functionalis layer is more vulnerable to vascular permutations 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 persisding evidence of necrosis, while 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. 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. Rapid reepithelialization 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. By day 5–6, the entire cavity is reepithelialized, 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, corresponding to peak estradiol levels in the circulation and maximal estrogen receptor concentration in the endometrium. This proliferation is marked by increased mitotic activity and increased nuclear DNA and cytoplasmic RNA synthesis, that 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. This proliferation is mainly in the functionalis 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 reexpanded 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. 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 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
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 17b-hydroxysteroid dehydrogenase and sulfotransferase, which convert estradiol to estrone sulfate (which is rapidly excreted from the cell). 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.

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 glycoprotein vacuoles in the glandular epithelium on cycle days 17–18. Giant 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 indiffident 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 production by the endometrium. An increase in capillary permeability is a consequence of this local increase in prostaglandins. 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 to 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 placation. 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 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.

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 assigns a key role to decidual cells in both the process of endometrial bleeding (menstruation) and the process of endometrial hemostasis (implantation and placentation). 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 functionals 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 life span of the corpus luteum is completed, and estrogren 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 following vascular sequence has been constructed from direct observations of theusus endometrium. 2 With shrinkage of height, blood flow within the spiral vessels diminishes, venous drainage is decreased, and vasodilatation ensues. Thereafter, the spiral arterioles undergo rhythmic vasocostriction and relaxation. Each successive spasm is more prolonged and profound, leading eventually to endometrial blanching.

Within the 24 hours immediately preceding menstruation, these reactions lead to endometrial ischemia and stasis. White cells migrate through capillary walls, at first remaining adjacent to vessels, but then extending throughout the stroma. During arteriole vasomotor changes, red blood cells escape into the interstitial space. Thrombin-platelet plugs also appear in superficial vessels. The prostaglandin content (PGF₂α and PGE₃) in the secretory endometrium reaches its highest levels at the time of menstruation. The vasocostruction and myometrial contractions associated with the menstrual events are believed to be significantly mediated by PGE₃ from glandular cells and the potent vasconstrictor, endothelin-1, derived from stromal decidual cells.

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 endometrial 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. 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.

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. The metalloproteinases include collagenases that degrade interstitial and basement membrane collagens, gelatins that further degrade collagens, and stromelysins that degrade fibronectin, laminin, and glycoproteins. The expression of metalloproteinases in human endometrium 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. Progestosterone withdrawal from endometrial cells induces metalloproteinase secretion, which is followed by the breakdown of cellular membranes and the dissolution of extracellular matrix.

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 TGF-β. 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. Thus, progesterone withdrawal can lead to endometrial breakdown through a mechanism that is independent of vascular events (specifically ischemia), a mechanism that involves cytokines. 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.

There is considerable evidence to support a major role for a cytokine, tumor necrosis factor-α (TNF-α) in menstruation. 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.

Eventually, considerable leakage occurs as a result of diapedesis, and finally, interstitial hemorrhage occurs due to breaks in superficial arterioles and capillaries. As ischemia and weakening progress, the continuous binding membrane is fragmented, and intercellular blood is extruded into the endometrial cavity. New blood flow within the 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. In particular, exerts an important restraining action on fibrinolysis and proteolytic activity.

With further tissue disorganization, the endometrium shrinks even more and coiled arterioles are buckled. Additional ischemic breakdown ensues with necrosis of cells and defects in vessels adding to the menstrual effluvium. 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 mm to 1.25 mm. Menstrual flow stops as a result of the combined effects of prolonged vasoconstriction, 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.

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). 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 (Chapter 15).

Dating the Endometrium

The postovulatory endometrium can be dated according to the histologic changes throughout a hypothetical 28-day menstrual cycle. These changes were described by Noyes, Hertig, and Rock in the lead article of the first volume of Fertility and Sterility in 1950. Dating of the endometrium is most accurately accomplished with biopsy specimens obtained 2–3 days before the onset of menses. This method continues to be the most accepted way to diagnose an inadequate luteal phase (endometrium inadequate to sustain a pregnancy because of deficient progesterone secretion by the corpus luteum).

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 by Rock et al. 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 2 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 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. 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.
loss can be achieved with treatment utilizing one of the nonsteroidal anti-inflammatory agents that inhibit prostaglandin synthesis. These agents are also effective in reducing menstrual blood flow in women with excessive menstrual bleeding due to alterations in prostaglandin production.

Because of its rapid metabolism, TGF-a is limited to autocrine and paracrine activity. Thromboxane is synthesized by uterine tissues. Gene expression for the thromboxane synthase and for the thromboxane receptor can be identified in endometrial stromal and epithelial cells. Thromboxane A2 is a potent vasoconstrictor and stimulator of smooth muscle cells.

Prostaglandin production decreases dramatically after implantation, suggesting the presence of an active mechanism for suppression. The production of prostaglandins requires estrogen support, but the increased production by secretory endometrium suggests progesterone enhancement, and acute withdrawal of progesterone promotes a further increase. Endometrial stromal cells produce prostacyclin and thromboxane in response to estrogen, a response that can be blocked by progestins. 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 F2α.

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. Thromboxane A2 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 in reducing menstrual blood flow in women with excessive menstrual bleeding due to alterations in prostaglandin production.
Fibronectin and laminin are extracellular matrix substances that are secreted by stromal cells of the endometrium in response to progesterone. 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. The expression of integrins appears to be regulated by cytokines and growth factors, not estrogen and progesterone.

Uteroglobin is a small protein produced in endometrial epithelial cells. The physiologic function of uteroglobin is uncertain. Uteroglobin, with high affinity, binds prostaglandins and may play a role in immunosuppression. Uteroglobin gene expression is stimulated by estrogen, and this response is enhanced by progesterone. Human decidua can secrete β-endorphin, yet another candidate for involvement in endometrial immunologic events, and its release is inhibited by both estrogens and glucocorticoids.

Endothelins are potent vasoconstrictor peptides produced in the vascular endothelial cells. The vasocostrictor activity of endothelin-1, present in the decidua, 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. 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 reepithelialization of the endometrium. Human decidual cells also synthesize and secrete endothelin-1, from where it may be transported into the amniotic fluid.

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. Vascular endothelial growth factor, a specific mitogen for endothelial cells, is abundantly expressed in human endometrium, reaching a peak that correlates with the maximal angiogenesis reached during the secretory phase. Angiogenesis is also influenced by many of the growth factors, and other substances such as fibronectin and prostaglandins. Fibroblast growth factor, in particular, is highly mitogenic for endometrial cells as well as endometrial stromal cells.

In all types of endometrial and myometrial cells, estrogen receptor expression reaches a maximum in the late follicular phase. The concentration is greatest in the glandular epithelium. During the early luteal phase, estrogen receptor expression declines, followed by an increase in the mid and late luteal phases. These changes reflect the cyclic changes in estradiol (which increases estrogen receptor expression) and progesterone (which decreases estrogen receptor expression).
membranes is in the direction of the amniotic cavity. The amniotic fluid concentration correlates with the decidual content, not maternal circulating levels. Amniotic fluid prolactin reaches peak levels in the first half of gestation (about 4000 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. It can be demonstrated that 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 contributes to the prevention of immunologic rejection of the conceptus. Prolactin can also function as an autocrine and paracrine growth factor in the uterus. 12

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. 13 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. 14

Prorenin (the inactive precursor of renin) is produced in decidua in response to IGF-1, insulin, endothelin, and relaxin. 15, 16 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. Large amounts of IGFBP-1 are present in amniotic fluid. The IGFBPs appear to be regulated by insulin, the IGFs, and relaxin. 17 Relaxin is related structurally to insulin and the IGFs. However, in contrast to insulin and IGF, it stimulates IGFBP-1 production in endometrial stromal cells. 18

IGFBP-1 begins to appear in midluteal 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 a-globulin. By the second trimester of pregnancy, high levels of IGFBP-1 are present in the amniotic fluid and the circulation, which 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. 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. 19 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 chorion laeve, villous trophoblast, and decidua are all sites of TGF-b production. 20 TGF-b can signal its own production; thus, TGF-b can be a messenger from fetal tissues to decidua. TGF-b is also believed to play a role in limiting trophoblastic invasion. 21 This may be accomplished by stimulating the production of plasminogen activator inhibitor and the factor that causes tissue inhibition of metalloproteinases. 22

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 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 4, 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 and contribute to the problems of infertility, recurrent pregnancy loss, and poor outcome in pregnancy (encountered in approximately 25% of women with uterine anomalies). 23-25 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. 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.

<table>
<thead>
<tr>
<th>Incidence of Müllerian Defects 57</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Fertile women</td>
</tr>
<tr>
<td>Infertile women</td>
</tr>
<tr>
<td>Women with recurrent miscarriages</td>
</tr>
<tr>
<td>Women with late miscarriages</td>
</tr>
<tr>
<td>and preterm deliveries</td>
</tr>
</tbody>
</table>

Anomalies can originate in 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 may 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 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 effluxus. 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).

<table>
<thead>
<tr>
<th>Distribution of Specific Anomalies 58</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicornuate uterus</td>
</tr>
<tr>
<td>Arcuate uterus</td>
</tr>
<tr>
<td>Incomplete septum</td>
</tr>
<tr>
<td>Uterus didelphys</td>
</tr>
<tr>
<td>Complete septum</td>
</tr>
<tr>
<td>Unicornuate uterus</td>
</tr>
</tbody>
</table>

Uterine anomalies can be organized into the following categories. 59 Each of these can be associated with obstructions that present during adolescence with amenorrhea and cyclic pain. 60
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 obstetrical complications (early spontaneous miscarriage, ectopic pregnancy, abnormal presentations, intrauterine growth retardation, and premature labor). 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 may be worthwhile. However, most rudimentary horns are asymptomatic because they are non-communicating, 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).

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. 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 malpresentations and premature labor, although many patients will have no reproductive difficulties.

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.

The Septate Uterus

Partial lack of resorption of the midline septum between the two müllerian ducts results in 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. The complete septate uterus is associated with a high risk of preterm labor and breech presentation. Outcomes are excellent with treatment by hysteroscopy. Posttreatment miscarriage rates are approximately 10% in contrast to the 90% pretreatment rates. A longitudinal vaginal septum usually does not have to be excised (unless dyspareunia is a problem). In some reports, the arcuate uterus had no adverse impact on reproductive outcome.

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 little success, and repeat surgery due to reappearance of obstruction is common. 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 31)

The Diethylstilbestrol-Associated Anomaly

We are still encountering women whose mothers were treated with high doses of estrogen during their pregnancies. Exposure to these high levels of estrogen during müllerian development caused a variety of anomalies, ranging from the hypoplastic "T" shaped uterus to irregular cavities with adhesions. 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. 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 has required surgical intervention, first laparotomy and then, more recently, laparoscopy. Today, vaginal ultrasonography and magnetic resonance imaging are highly accurate, and surgical intervention is usually not necessary. The hysterosalpingography is relatively inaccurate, and decisions should not be based upon hysterosalpingography alone. Congenital anomalies of the müllerian ducts are frequently accompanied by abnormalities in the urinary tract. Renal agenesis is often present on the same side as a müllerian defect.

Leiomyomata (Uterine Fibroids)

Uterine leiomyomas are benign neoplasms that arise from uterine smooth muscle. It is hypothesized that leiomyomas originate from somatic mutations in myometrial cells, resulting in progressive loss of growth regulation. 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. Different rates of growth can reflect the different cytogenetic 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; however, the familial inheritance of uterine myomas has not been well studied. It is not certain whether leiomyosarcomas arise independently or from leiomyomas. However, the incidence of leiomyosarcomas in patients with leiomyoma is very low (less than 1%).

If surgical specimens are serially sectioned, about 77% of women who come to hysterectomy will have myomas, many of which are occult. Overall, about 17% of hysterectomies are performed for myomas in the U.S. (44% in women 45–54 years old). The peak incidence for myomas requiring surgery occurs around age 45, approximately 8 cases per 1000 women each year. In the U.S., 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. The major symptoms associated with myomas are menorrhagia and the physical effects produced by large myomas.

Myomectomy (either abdominal myomectomy or hysteroscopic resection) is about 15%, with subsequent hysterectomy necessary in 1–5% of patients. In a series with long-term follow-up, the recurrence rate over 10 years reached 27%. Women who gave birth after myomectomy had a peak incidence for myomas requiring surgery occurs around age 45, approximately 8 cases per 1000 women each year. In the U.S., 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. The major symptoms associated with myomas are menorrhagia and the physical effects produced by large myomas.

Myomectomy will be encountered in about 1% of pregnant women. The risk of myoma is decreased with increasing parity and with increasing age at last term birth. 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 a decrease in BMI, 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.

The hormone sensitivity of leiomyomas is indicated by the following clinical observations. Leiomyomata 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. Aromatase gene and enzyme expression are present in significant levels in leiomyomata. Indeed, leiomyoma tissue is hypersensitive to estrogen and appears to have lost a regulatory influence that limits estrogen response. Endometrial hyperplasia is frequently observed at the margins of submucous myomas. 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. These facts indicate that progesterone stimulates mitotic activity in leiomyomas, but 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). Treatment with RU486, the progesterone antagonist, is associated with a reduction in leiomyomata size.

The bcl-2 protooncogene, presumably a cell survival gene, produces a protein that prevents apoptosis and promotes cell replication. Bcl-2 protein expression is increased in leiomyoma cells, and markedly increases in response to progesterone. 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. This further supports a key role for progestins in leiomyoma growth.

As in the normal uterus, the effects of estrogen and progestins on leiomyomata are mediated by growth factors. EGF is overexpressed in myomas, EGF receptors are present in leiomyomata, and GnRH agonist treatment (and hypogonadism) decreases EGF concentration in myomas (but not in normal myometrium). IGF-I and IGF-II and their receptors are abundant in myometrium, and actively overexpressed in leiomyomas. Leiomomas express more IGF-II than IGF-I, and to a lesser extent than myometrium, a situation that would enhance growth factor availability and activity in the tumor. Leiomomas cells express more parathyroid hormone-related protein (another growth factor) than normal myometrium. Like the endometrium and myometrium, leiomyomas secrete prolactin, and prolactin functions in the uterus as a growth factor. Even hematopoiesis is possible in a leiomyoma.

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

Reproductive Function and Leiomyomata

Leiomyomata are an infrequent cause of infertility, either by mechanical obstruction or distortion (and interference with implantation). 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, myometomy is usually followed by a prompt achievement of pregnancy in a high percentage of patients (usually within the first year). Submucous myomas are best treated by hysteroscopic resection. Preoperative visualization is important, and mapping of myomas by magnetic resonance imaging (MRI) is superior to ultrasonography (which is relatively inaccurate). It is difficult to distinguish between submucous myomas and endometrial polyps with ultrasonography. 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 short-term recurrence rate after myomectomy (either abdominal myomectomy or hysteroscopic resection) is about 15%, with subsequent hysterectomy necessary in 1–5% of patients. In a series with long-term follow-up, the recurrence rate over 10 years reached 27%. 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. 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. Intrauterine myomas, however, usually require surgery. Because of the rapid regrowth of myomas following cessation of GnRH agonist therapy, medical therapy for infertility is not recommended.

Myomas are present (diagnosed by ultrasonography) in about 1–2% of pregnancies. Most myomas do not grow during pregnancy. 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). So-called red degeneration of myomas is occasionally observed during late pregnancy, a condition due to central hemoragic 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. Surgery should be a last resort. The larger the myoma, the greater the risk of premature labor.

Medical Therapypf Leiomyomatap

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.

The short half-life of GnRH is due to rapid clevage 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 (inhbilitating 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 lutening 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 hypoestrogenic, 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 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 leiomyomata, the mean uterine size decreases 30–64% after 3–6 months of treatment. 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 agonist treatment. A decrease in operative blood loss can be achieved when the pretreatment uterus is as large or larger than a 15-week pregnancy. 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 values 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 submucous tumors. In addition to a decrease in myoma size, endometrial atrophy will improve visualization, and decreased vascularly will reduce blood loss.

Leiomyomatosis Peritonealis Dissemnata 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.

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.

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, vaginal dryness, joint and muscle stiffness, and depression. About 35% 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. On the other hand, Friedman and colleagues have suggested that maintaining the estradiol level in the early follicular phase range (30–50 pg/mL) can protect against osteoporosis and reduce hot flushes, but not allow the growth of myomas. The efficacy of this titration of response requires validation by clinical studies.

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.

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 denegregation and necrosis of submucous myomas. 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 growth. Most patients with leiomyosarcoma are women in the fifth and sixth decade, and about one-third of patients are postmenopausal. The incidence of leiomyosarcoma, even in patients with “rapidly growing leiomyomata,” is very low (less than 0.5%) and almost unheard of in premenopausal women.

Surgery should be a last resort. The larger the myoma, the greater the risk of premature labor.

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. 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 connective tissue disorders, 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. 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 16–20) effectively reduced uterine volume and maintained the reduced volume for 2 years (and avoided any loss in bone density). A daily 2.5 mg dose of ibutalene also prevents bone loss and inhibits vasomotor symptoms without reducing the therapeutic efficacy of GnRH agonist treatment.

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 leiomyomata tissue to progesterone agents, it makes sense to keep the dose of progestin...
Summary of Clinical Advantages With GnRH Agonist Treatment

**Reduction in menstrual blood loss.**

Improvement in amenorrhea prior to surgery. **Time for autologous blood donation.**

Less operative blood loss. **Less endometrial scarring.**

Hysterecetomy less likely. **More likely to allow laparoscopic technique.**

Possible conversion from abdominal to vaginal hysterectomy.

**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.

**Chapter References**

There are two major sites of action within the brain which 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, developments over the past two decades indicate that 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 which originate in the posterior pituitary that in turn receives its arterial supply from the inferior hypophyseal arteries.

**The Neurohormone Concept**

A considerable body of evidence indicates that 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 above, 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 hypothalamic control. Furthermore, cultures of anterior pituitary tissue release prolactin in the absence of hypothalamic tissue or extracts.

Neuroendocrine agents originating in the hypothalamus have positive stimulatory effects on growth hormone, thyroid-stimulating hormone (TSH), adrenocorticotropin 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 Besides being the principal regulator of ACTH secretion, 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. 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 apparent 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 being transported down the axon to the terminal for secretion. Small neuroendocrine peptides share common large precursor polypeptides, called polypeptides 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 is located on the short arm of chromosome 8. The precursor protein for GnRH contains (in the following order) a 23 amino acid signal sequence, the GnRH decapetide, a 3 amino acid proteolytic processing site, and a 56 amino acid sequence called GAP (GnRH-associated peptide). 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. 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 extratubular tissues (such as 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. 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 8p21-p11.2. Both genes produce a peptide with a signal sequence, a decapeptide, and a GAP. Human GnRH-I expression is highest outside the brain. An analysis of the evolution of GnRH indicates 3 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.

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. 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 lactotrophs of the anterior pituitary gland, in decidualized endometrium, and 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. In addition, 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, and therefore mutations in this gene can lead not only to absent secretion of growth hormone, prolactin, and TSH, but also an absence of their trophic cells in the pituitary; the result is significant hypopituitarism. 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.

The main function of prolactin in mammals is lactogenesis, while 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. The Pit-1, the pituitary specific transcription factor, therefore, appears to be highly conserved among species.

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 under the inhibitory control of hypothalamic dopamine released into the portal circulation. 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, D1 and D2. The D2 type is the predominant receptor in the anterior pituitary gland. The structure and function of the dopamine receptors are of the G protein system as described in Chapter 2. Binding of dopamine to the receptor leads to suppression of adenylate cyclaase 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 D2 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.

The secretion of prolactin is inhibited and stimulated by the association and dissociation of dopamine from its receptors. 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.

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
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 1000–3000 GnRH-producing cells can be found in the arcuate nucleus of the hypothalamus. 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 nerve terminals, a cranial nerve that projects from the nose to the septal-preoptic nuclei in the brain. 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. 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 a protein (homologous to members of the fibroinectin family) responsible for cell adhesion and protease inhibition, functions necessary for neuronal migration. Like olfactory epithelial cells in the nasal cavity, GnRH neurons have cilia. 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 shorted the cycles of recipient women, and compounds from the luteal phase had the opposite effects. This may be one mechanism by which women who are together much of the time often exhibit a synchrony in menstrual cycle timing.

In primates, the primary network of GnRH cell bodies is located within the medial basal hypothalamus. 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. 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. 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 upon 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 upon 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. GnRH neurons lack estradiol receptors; therefore, steroid hormone regulation is believed to be mediated through this collection of neurotransmitters.

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, but dopamine itself functions as a key neurotransmitter in the hypothalamus and the pituitary.

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. Even pituitary hormone gene transcription is sensitive to the pulsatile nature of GnRH release.

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 shorted the cycles of recipient women, and compounds from the luteal phase had the opposite effects. This may be one mechanism by which women who are together much of the time often exhibit a synchrony in menstrual cycle timing.

In primates, the primary network of GnRH cell bodies is located within the medial basal hypothalamus. 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. 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. 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 upon 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 upon 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. GnRH neurons lack estradiol receptors; therefore, steroid hormone regulation is believed to be mediated through this collection of neurotransmitters.

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, but dopamine itself functions as a key neurotransmitter in the hypothalamus and the pituitary.

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. Even pituitary hormone gene transcription is sensitive to the pulsatile nature of GnRH release.
Experimental manipulations have indicated that the critical range of GnRH pulsatile secretion is rather narrow. The administration (to monkeys) of 1 mg GnRH per minute for 6 minutes every hour (1 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. 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.

Timing of GnRH Pulses

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

**LH pulse mean amplitude:**
- Early follicular phase: 6.5 IU/L
- Late follicular phase: 7.2 IU/L
- Early luteal phase: 15.0 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 to 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. Despite the handicap of the long half-life, it has been ascertained that FSH secretion is correlated with LH secretion.

The anterior pituitary gland also appears to have a pulsatile pattern of its own. Although pulses of significant amplitude are linked to GnRH, small amplitude pulses of high frequency represent spontaneous secretion (at least as demonstrated in isolated pituitary glands in vitro). It is not known whether this has any importance physiologically, and at the present time, the pituitary secretory pattern is thought to reflect GnRH.
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. The administration of dopamine by intravenous infusion to men and women is associated with a suppression of circulating prolactin and gonadotropin levels. 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.

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.

Neuropeptide Y. The secretion and gene expression of neuropeptide Y in hypothalamic neurons is regulated by gonadal steroids. Neuropeptide Y stimulates pulsatile release of GnRH and in the pituitary potentiates gonadotropin response to GnRH. 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 and increased amounts have been measured in cerebrospinal fluid of women with anorexia and bulimia nervosa, it has been proposed that neuropeptide Y is at least one link between nutrition and reproductive function.

Pituitary Gonadotropin Secretion

The gene for the a subunit of the gonadotropins is expressed in both the pituitary and placenta. The b-subunit for human chorionic gonadotropin (HCG) is expressed in the placenta but only minimally (and with alterations in structure) in the pituitary, while the LH b-subunit, as expected, is expressed in the pituitary but not significantly in the placenta. 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.

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 activity (Chapter 2). 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, both protein kinase and calmodulin are mediators of GnRH action. The GnRH receptor, a member of the G protein family, is encoded by a gene on chromosome 4q13—14q21.1. GnRH receptors are regulated by many agents, including GnRH itself, inhibin, activin, and the sex steroids. A decreased gonadotropin response to continued excessive GnRH stimulation is not due to a loss of GnRH receptors alone but includes desensitization and uncoupling of the receptors (discussed in Chapter 2).

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, while
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 upon 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.

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, and many others. 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-b family. 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. Inhibin selectively inhibits FSH, but not LH, secretion. Indeed, while suppressing FSH synthesis, inhibin may enhance LH activity. Cells actively synthesizing LH respond to inhibin by increasing GnRH receptor number; FSH dominant cells are suppressed by inhibin. Inhibin has little or no effect on growth hormone, ACTH, and prolactin production.

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. Activin augments the secretion of FSH and inhibits prolactin, ACTH, and growth hormone responses. Activin increases pituitary response to GnRH, probably by enhancing GnRH receptor formation. The effects of activin are blocked by inhibin and follistatin. The roles for inhibin and activin in regulating the events of the menstrual cycle are discussed in Chapter 6.

The Two Forms of Inhibin

- Inhibin-A: Alpha-Beta_A
- Inhibin-B: Alpha-Beta_B

The Three Forms of Activin

- Activin-A: Beta_A-Beta_A
- Activin-AB: Beta_A-Beta_B
- Activin-B: Beta_B-Beta_B

Follistatin is a peptide secreted by a variety of pituitary cells, including the gonadotropes. This peptide has also been called FSH-suppressing protein because of its main action: inhibition of FSH synthesis and secretion. Inhibin selectively inhibits FSH, but not LH, secretion. Indeed, while suppressing FSH synthesis, inhibin may enhance LH activity. Cells actively synthesizing LH respond to inhibin by increasing GnRH receptor number; FSH dominant cells are suppressed by inhibin. Inhibin has little or no effect on growth hormone, ACTH, and prolactin production.

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. This relationship may contribute 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.

The Endogenous Opiates

The most fascinating peptide group is the endogenous opioid peptide family. b-Lipotropin is a 91 amino acid molecule which 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 physiological 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. All opiates derive from one of 3 precursor peptides.

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

![Image of the endogenous opioid peptide family](image_url)
is in the pituitary gland.

Proopiomelanocortin is split into 2 fragments, an ACTH intermediate fragment and b-lipotropin. b-Lipotropin has no opioid activity but is broken down in a series of steps to b-melanocyte-stimulating hormone (b-MSH), enkephalin, and a-, g-, and b-endorphins. Melanocystimulating 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 a- and g-endorphins are as active as morphine on a molar basis, while b-endorphin is 5–10 times more potent. In the adult pituitary gland, the major products are ACTH and b-lipotropin, with only small amounts of endorphin. Thus, ACTH and b-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 b-MSH. In the placenta and adrenal medulla, POMC processing yields a-MSH-like and b-endorphin peptides. b-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 b-endorphin and a-MSH in the region of the arcuate nucleus and the ventromedial nucleus. The pituitary system is a system for secretion into the circulation while the hypothalamic system allows for distribution via axons to regulate other brain regions and the pituitary gland.

b-Endorphin is appropriately considered a neurotransmitter, a neurohormone, and a neuromodulator. b-Endorphin influences a variety of hypothalamic functions, including regulation of reproduction, temperature, 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. 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 a-neoendorphin, b-neoendorphin, and leumorphin. The last 13 amino acids of leumorphin constitute another opioid peptide, rimorphin. The prodynorphin protein probably function in a fashion similar to endorphin.

It is simpler to say that there are 3 classes of opiates: enkephalins, 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 pharmacological 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. 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. 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. 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. 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. Nevertheless, the evidence overall indicates that endogenous opiates exert an inhibiting influence over GnRH secretion.

The inhibiting tone of endogenous opiates is reduced at the time of the ovulatory surge, allowing a release from suppression. This is probably a response to estrogen, specifically an estrogen-induced decrease in opioid receptor binding and opioid release.

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 b-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 neurotransmitter.

Because a-MSH counteracts the effects of b-endorphin, posttranslational processing of POMC can affect hypothalamic-pituitary function by altering the amounts of a-MSH and b-endorphin. 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, while endogenous opioid inhibition does not seem to play a causal role in delayed puberty or hereditary problems such as Kallmann’s syndrome. Treatment of patients with hypothalamic amenorrhea (suppressed GnRH pulsatile secretion) with a drug (naltrexone) which blocks opioid receptors restores normal function
(ovulation and pregnancy). 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) directly inhibits hypothalamic GnRH secretion, both directly and by augmenting endogenous opioid secretion. Women with hypothalamic amenorrhea demonstrate hypercortisolism, suggesting that this could be the pathway by which stress interrupts reproductive function. Mathematical analysis of the associations among FSH, LH, b-endorphin and cortisol pulses support the existence of significant functional coupling between the neuroregulatory systems that control the gonadal and adrenal axes. 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.

Cumming concludes that most studies indicate an exercise-induced increase in endogenous opiates, but a significant impact on mood remains to be substantiated. He notes 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 b-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 physiological 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.

Every pituitary hormone appears to be modulated by opiates. Physiologic effects are important with ACTH, gonadotropins, and possibly vesopressin. 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. 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 b-endorphin shows significant changes during the menstrual cycle, reaching highest levels just before ovulation.

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. To be specific, catecholestrogens can inhibit tyrosine hydroxylase (which would decrease catecholamines) and compete for catechol-α-methyltransferase (which would increase catecholamines). Since 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. The normal physiology and pathophysiology of the menstrual cycle, at least in terms of central control, can be explained by mechanisms which affect the pulsatile secretion of GnRH. The pulses of GnRH appear to be 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 may 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. Thousands of GnRH analogues have been produced. 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, while 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 utilized for the treatment of endometriosis, uterine leiomyomas, precocious puberty, or the prevention of menstrual bleeding in special clinical situations (e.g., in thrombocytopenic patients). Various tumors contain receptors for GnRH, such as breast, pancreatic, and ovarian, and therefore, there exists a potential for treatment.

GnRH antagonists are synthesized by 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. The early products either lacked potency or were associated with undesirable side effects due to histamine release. New analogues continue to be developed and tested, aimed toward the control of fertility. The combination of a GnRH antagonist and testosterone holds promise as a male contraceptive agent.
The GnRH analogues cannot escape destruction if administered orally. Higher doses administered subcutaneously can achieve nearly equal effects as 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 injections of biodegradable microspheres. With the nasal route, absorption enhancers have to be added to increase bioavailability; these agents produce considerable nasal irritation. Goserelin consists of a small biodegradable cylinder which is inserted subcutaneously and monthly using a prepackaged syringe. The depot formulations of GnRH agonists are administered intramuscularly and monthly.

<table>
<thead>
<tr>
<th>GnRH Agonists in Clinical Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
</tr>
<tr>
<td>Native GnRH</td>
</tr>
<tr>
<td>Leuprolide</td>
</tr>
<tr>
<td>Buserelin</td>
</tr>
<tr>
<td>Nafarelin</td>
</tr>
<tr>
<td>Histrelin</td>
</tr>
<tr>
<td>Goserelin</td>
</tr>
<tr>
<td>Deslorelin</td>
</tr>
<tr>
<td>Tryptorelin</td>
</tr>
</tbody>
</table>

Tanyocytes

A significant pathway for hypothalamic influence may be via the cerebrospinal fluid (CSF). Tanyocytes 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. Tanyocytes change morphologically in response to steroids and exhibit morphological 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. Both oxytocin and vasopressin consist of 9 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. 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. 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.

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 which 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, in the case of vasopressin, ACTH secretion, and in the case of oxytocin, gonadotropin secretion.

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. 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 to posterior pituitary release.

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. Oxytocin neurons and vasopressin neurons have been demonstrated in the rat to contain the estrogen receptor-beta. 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 may 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 involve the regulation of osmolality and blood volume. Vasopressin is a powerful vasconstrictor 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, secretory 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 as well as 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 3 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. In the male, release of oxytocin during coitus may contribute to sperm transport during ejaculation.

Using sensitive assays, an increase in maternal levels of oxytocin can be detected prior to parturition, occurring at first only at night. 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. Part of the contribution of oxytocin to parturition is the stimulation of prostaglandin synthesis in decidua and myometrium. Cervical dilatation 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, chorion, and significantly, in the decidua. 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 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. The mechanism for the increase is unknown, but it is likely 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 3rd, 4th, and 5th thoracic nerves to the spinal cord 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 Ovation

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. In the case of LH, there existed both a negative inhibitory feedback relationship with estradiol and a positive stimulatory feedback with high levels of estradiol. The feedback centers were located in the hypothalamus, and they 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 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.

In the primate, the “center” for the midcycle surge of gonadotropins has moved from the hypothalamus to the pituitary. Experiments in the monkey have demonstrated that GnRH, originating in the hypothalamus, plays a permissive and supportive role. Its pulsatile secretion is an important prerequisite for normal pituitary function, 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 (MBH) is either destroyed or the hypothalamus is surgically separated from the pituitary. In a typical (and now classic) experiment, lesion of the MBH by radiofrequency waves was followed by loss of LH levels as the source of GnRH was eliminated. 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, clearly actions that 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, then a second and sustained increase at 225–240 minutes. This biphasic response suggests the presence of two functional pools of pituitary gonadotropins. The readily releasable pool (secretion) produces the initial response, and the later response is dependent upon 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 in order 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. Estrogen itself is capable of increasing the number of GnRH receptors. 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 now 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. The negative feedback of estrogen operates through an uncertain, but different, system.

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.

GnRH is increased in the peripheral blood of women and the portal blood of monkeys at midcycle. While this increase may not be absolutely necessary (as demonstrated in the monkey experiments), studies do indicate that activity is occurring in both the hypothalamus and the pituitary. 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. The upstream region of the LH b-subunit gene (in the rat) binds the estrogen receptor, providing a means for direct steroid hormone modulation in the pituitary. The human GnRH gene contains a hormone responsive element that binds estrogen and its receptor. However, studies have failed to detect the presence of estrogen receptors in GnRH neurons, and this hormone responsive element may be regulated by other substances, or other estrogen responsive neurons may synapse with GnRH cell bodies. Another possibility is that GnRH neurons contain estrogen receptor-beta, and the previous immunoreactive studies have been directed to estrogen receptor-alpha. Nevertheless, 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. Certainly, the presence of GnRH is essential; the administration of a GnRH antagonist to women at midcycle prevents the LH surge.
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. 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.

Another aspect of gonadotropin secretion is clinically important. A disparity exists between the quantity of LH measured during the midcycle surge as determined by immunoassay and bioassay. More LH is secreted at midcycle in a molecular form with greater biological activity. 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 upon pulsatile stimulation by GnRH.

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 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. 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, and 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. First, there definitely is a central action to decrease GnRH. 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 addition, progesterone can also block estrogen-induced responses to GnRH at the pituitary level. In contrast, the facilitatory action of low levels of progesterone is exerted only at the pituitary on the response to GnRH.

**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. Low levels of estrogen enhance FSH and LH synthesis and storage, have little effect on LH secretion, and inhibit FSH secretion.
4. High levels of estrogen induce the LH surge at midcycle, and high steady levels of estrogen lead to sustained elevated LH secretion.
5. 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.
6. 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 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 which allows the pineal gland to be an active neuroendocrine organ that responds to photic and hormonal stimuli and exhibits circadian rhythms.

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 adenylyl 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. Thus, melatonin synthesis is controlled by norepinephrine stimulation of adenylyl 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 cerebellum. 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.

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 photoperiod. Short days and short light result in gonadal atrophy, and this is the major mechanism governing seasonal breeding. In humans, melatonin secretion increases after darkness and 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 to 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.

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 as 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 tanyocyte 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. In humans, melatonin blood levels are highest in the first year of life (with highest levels at night), then decrease with age, eventually releasing, some claim, the suppression of GnRH prior to puberty. This hypothesis is challenged by the association of blindness in human females with an age of menarche that is earlier than normal. Furthermore, pinealectomy in monkeys does not affect puberty.

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 5a-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. 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. 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. 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 women with anorexia nervosa.

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 among roommates and among closest friends occurred in the first 4 months of residency in a dormitory of a women's college. 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. However, efforts to replicate these results have not always been successful.

Melatonin is available in 1 and 5 mg doses that produce blood levels that are 10 to 100 times higher than normal nighttime peaks. 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 elusive.

**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 4th and 5th 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. 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 being 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 23 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 masculinization differentiate to be maintained by the fetal pituitary gonadotropins. Decreased testosterone levels in late gestation probably reflect the decrease in in utero gonadotropins. The fetal generation of Leydig cells somehow avoids down-regulation and decreased steroidogenesis.

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 probably due to testicular testosterone and inhibit production. In infancy, the postnatal FSH rise is more marked and more sustained in females, while LH values are not as high. This activity is accompanied by inhibin levels comparable to the low range observed during the follicular phase of the menstrual cycle. 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. This childhood period is characterized by low levels of gonadotropins in the pituitary and in the blood, little response of the pituitary to
The precise signal that initiates the events of puberty is unknown. In girls, the first steroids to rise in the blood are dehydroepiandrosterone (DHA) and its sulfate (DHAS) 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 10–12 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 in the maturational change must be considered. However, there is no evidence to suggest that the adrenal steroids are necessary for the proper timing of puberty, and adrenarche appears to be independent, not controlled by the same mechanism that turns on the gonads. 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.

Prior to puberty, gonadotropin levels are low but still associated with pulses (although quite irregular). The clinical onset of puberty is preceded by an increase in pulse frequency, amplitude, and regularity, especially during the early night. 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 (such as in individuals with worsening and improving anorexia nervosa). FSH levels plateau by midpuberty, while LH and estradiol levels continue to rise until late puberty. Biologically active LH has been found to rise proportionately more than immuno-reactive LH with the onset of puberty.

The rise of gonadotropins at puberty appears to be independent of the gonads in that 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. 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 halted. In a 10-year prospective study of middle class contemporary American girls, the mean age of menarche was 12.83 with a range of 9.14–17.70 years. The age of onset of puberty is variable and influenced by genetic factors, socioeconomic conditions, and general health. The earlier menarche today compared to 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 (45 kg) and percent of body fat (17%). 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 period 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 steroids and therefore suppressed. The changes at puberty are due to a gradually increasing gonadotropin secretion that takes place because of a decrease in the sensitivity of the hypothalamic centers to the negative-inhibitory action of gonadal steroids. This can be pictured as a slowly rising set point of decreased sensitivity, resulting 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 accurate 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.

Negative feedback of steroids, however, cannot be the sole explanation for the low gonadotropin levels in children. Agonadal children show the same decline in gonadotropins from age 2 to 6 as do normal children. 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 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.

Chapter References

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 new understanding. The hormone changes, correlated with the morphologic and autocrine/paracrine 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 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 to emerge and pursue a path to ovulation. In addition, maintenance of this increase in FSH for a critical duration of time is essential. 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 period of 6–7 million in both ovaries. The primordial follicle is nongrowing 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 irrevocably 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. 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. The early growth of follicles occurs over the timespan of several menstrual cycles, but the oocyte is one of a cohort recruited at the time of the luteal-follicular transition. 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. 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. 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.

“Apoptosis” is derived from Greek and means falling off, like leaves from a tree.

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 may be 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 permit the exchange of nutrients, ions, and regulatory molecules. Thus, the gap junctions serve as the pathway for nutritional, metabolic, and signal interchange between the granulosa cells and the oocyte. The process of follicular growth is influenced by factors derived from the oocyte. Mice that are genetically deficient in growth differentiation factor-9 (GDF-9), a peptide synthesized only in the oocyte are infertile because follicular development cannot proceed beyond the primordial follicle stage.
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. 5

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. 11 12 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 growth. 13

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 60 days of development) responds to a hormonal change and is propelled to grow. In young women, this cohort numbers 3–11 per ovary. 14 The decline in luteal phase steroidogenesis and inhibit A secretion allows a rise in FSH, beginning a few days before menses. 15 16 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. 17

**Follicular Growth and Development Based on Nonprimate and Primate Data.** The following discussion of events, which mark the growth and development of the ovarian follicle from the preantral stage to ovulation, is based on a formulation that assigns a key role to estradiol function while recognizing the importance of local autocrine/paracrine factors. 18

**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 thecal 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 3 precursor cells. 19

The granulosa cells of the preantral follicle have the ability to synthesize all 3 classes of steroids; however, significantly more estrogens than either androgens or progesterones 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. 20

Specific receptors for FSH are not detected on granulosa cells until the preantral stage, 21 and the preantral follicle requires the presence of FSH in order to aromatize androgens and generate its own estrogenic microenvironment. 22 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. 23 This action of FSH is modulated by growth factors. 24 FSH receptors quickly reach a concentration of approximately 1500 receptors per granulosa cell. 25

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-a, 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. 26 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. 27 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. 28 These androgens cannot be converted to estrogen and, in fact, inhibit aromatase activity. 29 They also inhibit FSH induction of LH receptor formation, another essential step in follicular development. 30
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. 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.

### 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.

In the presence of FSH, estrogen becomes the dominant substance in the follicular fluid. Conversely, in the absence of FSH, androgens predominate. 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. 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.

### 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. Theca interstitial cells, located in the theca interna, have approximately 20,000 LH receptors in their cell membranes. In response to LH, thecal 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 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. 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 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 lipoproteins are not internalized, but rather, the cholesteryl esters are extracted from the lipoproteins at the cell surface and then transferred into the cell.

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. 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 thecal cells and P450arom only in granulosa cells is impressive evidence confirming the two-cell, two-gonadotropin explanation for estrogen production.

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. Follicles developed (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 stereoidogenesis was impossible without the presence of LH to produce thecal production of androgen substrate. This same response has been observed in experiments that use a GnRH antagonist to produce LH-deficient monkeys and then the administration of recombinant, pure human FSH. These results indicate that only FSH is required for folliculogenesis, and that in the primate, autocrine/paracrine peptides have replaced estrogen in the important intraovarian role of modulating gonadotropin response.

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. 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 (in the nonprimate model), 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, is a process called apoptosis, programmed cell death. This process is heralded by alterations in mRNAs required for cell proteins which maintain follicle integrity. 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. 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.

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 mid/follicular phase and preceding the increase in diameter that marks the physical emergence of the dominant follicle. 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.

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 (the nonprimate model) or because of local autocrine/paracrine peptides (as will be described below in the primate model). 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. 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 thecal vascularity. By day 9, thecal vascularity in the dominant follicle is twice that of other antral follicles. 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 second development points when proliferation of capillaries is important: the emerging dominant follicle and the early corpus luteum.

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 either estrogen (the nonprimate), or local autocrine/paracrine peptides (primate), serves as the chief coordinator.

In the nonprimate model, 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. 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-induced granulosa cells, but the primary mechanism utilizes FSH stimulation and estrogen enhancement. The role for estrogen goes beyond synergism and enhancement; it is obligatory. Inhibition of estrogen synthesis prevents FSH-stimulated increases in LH receptors.

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. In addition, an increase in GnRH accompanies the LH surge, indicating that estrogen positive feedback operates at both pituitary and hypothalamic sites.

Estrogen exerts its inhibitory effects in both the hypothalamus and the anterior pituitary, decreasing both GnRH pulsatile secretion and GnRH pituitary response. Progesterone also operates in two sites. Its inhibitory action is at the hypothalamic level, and, like estrogen, its positive action is directly
The secretions of FSH is very sensitive to the negative inhibitory effects of estrogen even at low levels. At higher levels, estrogen combines with inhibin for 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, estradiol secretion is not sufficient to achieve a positive feedback is more than 200 pg/mL, and this concentration must be sustained for approximately 50 hours. This level of estrogen essentially never occurs until the dominant follicle has reached a diameter of 15 mm. 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 (mean values below) is probably the consequence of steroid feedback on both hypothalamus and anterior pituitary.

### LH Pulse Frequency:

<table>
<thead>
<tr>
<th>Phase</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early follicular phase</td>
<td>~90 minutes.</td>
</tr>
<tr>
<td>Late follicular phase</td>
<td>~60–70 minutes.</td>
</tr>
<tr>
<td>Early luteal phase</td>
<td>~100 minutes.</td>
</tr>
<tr>
<td>Late luteal phase</td>
<td>~200 minutes.</td>
</tr>
</tbody>
</table>

### LH Pulse Amplitude:

<table>
<thead>
<tr>
<th>Phase</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early follicular phase</td>
<td>~6.5 I.U.</td>
</tr>
<tr>
<td>Midfollicular phase</td>
<td>~5.0 I.U.</td>
</tr>
<tr>
<td>Late follicular phase</td>
<td>~7.2 I.U.</td>
</tr>
<tr>
<td>Early luteal phase</td>
<td>~15.0 I.U.</td>
</tr>
<tr>
<td>Midluteal phase</td>
<td>~12.2 I.U.</td>
</tr>
<tr>
<td>Late luteal phase</td>
<td>~8.0 I.U.</td>
</tr>
</tbody>
</table>

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 more frequently 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. 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.

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. Both estradiol and progesterone are required to achieve the low, suppressed secretory pattern of GnRH during the luteal phase. The studies suggest that steroids influence the hypothalamic release of GnRH for frequency changes and the pituitary for action on amplitude of the gonadotropins. 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. 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. 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. 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. 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, both increase the quantity and the quality (the bioactivity) of FSH and LH.

There is a diurnal rhythm in FSH and LH secretion. 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. 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, probably 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. Initial efforts to understand the role of inhibin in physiology were hampered by imprecise assays and the fact that multiple forms are present in the circulation, including free subunits.

### The 2 Forms of Inhibin:

<table>
<thead>
<tr>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibin A</td>
<td>Alpha-Beta&lt;sub&gt;α&lt;/sub&gt;</td>
</tr>
<tr>
<td>Inhibin B</td>
<td>Alpha-Beta&lt;sub&gt;β&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

FSH stimulates the secretion of inhibin from granulosa cells and, in turn, is suppressed by inhibin — a reciprocal relationship. 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. 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 action of GnRH lends some support for an endogenous ovarian GnRH-like substance (which is found in follicular fluid) and which is involved in inhibin production.
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 min 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. 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.

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. The circulating levels of inhibin A rise in the late follicular phase to reach a peak level at the midluteal phase. 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 is a hormone that is related to inhibin but has an opposite action (the stimulation of FSH release and GnRH receptor number). This peptide contains two subunits that are identical to the beta-subunits of inhibins A and B. Thus, when each of the beta-subunits of the inhibins is combined with an alpha-subunit, the resulting molecule, inhibin A or B, inhibits the release of FSH. If the beta-subunits are paired together, the molecule stimulates the release of FSH. Each activin subunit is encoded by a distinct gene. 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. Another important member of this family is the antimitelhormone, as well as a protein active during insect embryogenesis, and a protein active in frog embryos.

The 3 Forms of Activin:

- **Activin A:** Betaα
- **Activin AB:** BetaαBetaβ
- **Activin B:** Betaβ

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. Considerable evidence derived from human cells exists to indicate that inhibin and activin act directly on theca cells to regulate androgen synthesis. 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. This receptor heterogeneity allows the many different responses elicted 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.

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.

In the anterior pituitary the secretions of inhibin and 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. Follistatin is a peptide secreted by a variety of pituitary cells, including the gonadotropes. 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, probably by binding to activin and in that fashion decreasing the activity of activin. 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. 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. 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 thecal and granulosal responses to the gonadotropins.

The inhibin-activin family of peptides (also including antimitelhormone 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. The mice that are homozygous and lack inhibin are susceptible to the development of gonadal stromal tumors which 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 somatotropins) are peptides that have structural and functional similarity to insulin and mediate growth hormone action. Insulin-like growth factor-I (IGF-I) and insulin-like growth factor-II (IGF-II) are single chain polypeptides containing 3 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...
Insulin-like Growth Factor Binding Proteins. There are 6 known nonglycosylated peptides that function as IGF binding proteins, IGFBP-1 to IGFBP-6. 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. In human ovaries, IGF-II is highly expressed in both thecal cells and granulosa cells; however, the level is highest in the granulosa and increases with growth of the follicle. IGF-II is also synthesized by luteinized granulosa and appears to function locally in an autocrine fashion. These findings indicate that IGF-II is the primary IGF in the human ovary. Nevertheless, IGF-I is still a significant product of human thecal cells.

In studies with human ovarian tissue, IGF-II is highly expressed in both thecal cells and granulosa cells; however, the level is highest in the granulosa and increases with growth of the follicle. IGF-II is also synthesized by luteinized granulosa and appears to function locally in an autocrine fashion. These findings indicate that IGF-II is the primary IGF in the human ovary. Nevertheless, IGF-I is still a significant product of human thecal cells.

Human thecal cells express mRNA transcripts which encode receptors for both IGF-I and insulin. 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.
IGF-II stimulates granulosa cell proliferation, aromatase activity, and progesterone synthesis. IGF-II is produced in thecal cells, granulosa cells, and luteinized granulosa cells. In the pig and rat, the primary IGF is IGF-I. The most abundant IGF in human follicles is IGF-II.

Ishikawa et al. (1993) have shown that IGF-I and IGF-II are present in human follicular fluid. In the dominant follicle, the levels of IGF-I and IGF-II are higher than in the preovulatory follicle. The ratio of IGF-II to IGF-I is higher in the dominant follicle, and this suggests that IGF-II is the important factor in the human dominant follicle.
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. Follicular levels of ACTH and b-endorphin remain constant throughout the cycle, but b-endorphin levels peak just before ovulation. In addition, enkephalin is present in relatively unchanged concentrations. The corticotropin-releasing hormone (CRH) system is present in thecal cells, but not in granulosa cells, complete with CRH, the CRH receptor, and CRH-binding protein. CRH inhibits LH-stimulated androgen production in thecal cells, apparently by suppressing P450c17 gene expression.

Antimüllerian hormone is produced by granulosa cells and may play a role in oocyte maturation (it inhibits oocyte meiosis) and follicular development. Antimüllerian hormone directly inhibits proliferation of granulosa and luteal cells, as well as epidermal growth factor-stimulated proliferation.

Follicular fluid prevents resumption of meiosis until the preovulatory LH surge either emerges 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. 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-ell, 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 day cycle 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.
6. FSH induces the appearance of LH receptors on granulosa cells.
7. In immature granulosa cells, activin augments all FSH activities, especially aromatase activity (estrogen production).
8. LH levels rise steadily during the late follicular phase, stimulating androgen production in the theca.
9. Inhibin B, secreted by the granulosa cells in response to FSH, directly suppresses pituitary FSH secretion.
10. As the follicle grows, activin production decreases and inhibin production increases. In addition, follistatin levels increase in follicular fluid with increasing growth of the follicle, a mechanism for decreasing activin activity.

Follicular Growth and Development in the Primate Ovary

Concern that the story for ovarian follicular growth and development might be different in the primate originated with the failure to find estrogen receptors in any of the significant ovarian compartments in the monkey: follicles, stromal tissue, interstitial tissue, or corpora lutea. The importance of this finding in the monkey has been challenged by the discovery that human granulosa cells contain only mRNA for estrogen receptor-beta. However, in further monkey experiments, no reduction in total number or size of follicles resulted when estradiol production was effectively suppressed by treatment with an inhibitor of 3b-hydroxysteroid dehydrogenase. 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).

A reduced role for estrogen is further supported by the response of women with a deficiency in women with deficiency in gonadotropins to treatment with recombinant (pure) FSH. 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 thecal 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. These results indicate that only FSH is required for folliculogenesis, and that in the primate, autocrine/paracrine peptides have replaced estrogen in the important role of modulating granulotroph response. Consider the following actions that have been documented in primate ovaries:

1. Inhibit and activate regulate androgen synthesis in human theca cells. Inhibit enhances and activin suppresses the stimulatory action of LH and/or IGF-I, and inhibit can overcome the inhibitory action of activin on theca cells.
2. In immature granulosa cells, activin augments all FSH activities, especially aromatase activity (estrogen production).
3. In luteinizing granulosa cells, activin has direct mitogenic activity and suppresses steroidogenesis in response to LH, while inhibit has no effect on LH-dependent aromatase in mature granulosa cells.
4. In the follicular phase, granulosa production of inhibit is under the control of FSH, but during the late follicular phase a change occurs, culminating in LH control of luteal synthesis of inhibit.
5. As the follicle grows, activin production decreases and inhibit production increases.

These actions may 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 thecal androgen synthesis. In the late follicular phase, increased production of inhibit 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).

The right concentration of androgens in granulosa cells promotes aromatase activity and inhibin production and, in turn, inhibin promotes LH stimulation of thecal androgen synthesis. With development of the follicle, inhibin expression 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.

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. 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. In addition, the IGFs in the ovary may not be growth hormone-dependent, and therefore IGF-I and/or IGF-II may also be important in primate ovarian follicles.

Summary of Events in the Primate Ovarian Follicle

1. FSH stimulates inhibin and activin production by granulosa cells.
2. Activin augments FSH activities: FSH receptor expression, aromatization, inhibin/activin production, and LH receptor expression.
3. Inhibin enhances LH stimulation of androgen synthesis in the theca to provide substrate for aromatization to estrogen in the granulosa.
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.
5. With the appearance of LH receptors, inhibin production is maintained as it comes under control of LH.
6. All functions are modulated by a host 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. The onset of the LH surge occurs when the peak levels of estradiol are achieved. 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). 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. 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.

Progesterone receptors begin to appear in the granulosa cells of the dominant follicle in the periovulatory period. The traditional view has been that progesterone receptors are expressed in response to estrogen through an estrogen-receptor mediated mechanism. This is not the case with the primate ovarian follicle. Experimental data in the monkey provide excellent evidence that LH stimulates progesterone receptor expression in granulosa cells. 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.

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. 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.
surge. Appropriately low levels of progesterone derived from the maturing follicle contribute to the precise synchronization of the midcycle surge.

In addition to its facilitatory action on LH, progesterone at midcycle is significantly responsible for the FSH surge. 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. Nevertheless, blockade of midcycle progesterone synthesis or activity in the monkey impaired the ovulatory process and luteinization. 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 P450cocc and P450cct, 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 P450cct. Other luteinized theca cells retain P450cct 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. There is a 15% increase in androstenedione and a 20% increase in testosterone. 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. The coital frequency of married couples has also been noted to increase at the time of ovulation. 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 thecaal 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. The onset of the LH surge appears to be the most reliable indicator of impending ovulation, occurring 34–36 hours prior to follicle rupture. A threshold of LH concentration must be maintained for 14–27 hours in order for full maturation of the oocyte to occur. Usually the LH surge lasts 48–50 hours.

Because of the careful timing involved in in vitro fertilization programs, we have available some interesting data. 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. 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.

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. 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.

The LH surge initiates the continuation 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, 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. LH-induced cyclic AMP activity overcomes the local inhibitory action of oocyte maturation inhibitor (OMI) and luteinization inhibitor (LI). LI may be endothelin-1, a product of vascular endothelial cells. 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. Usually the LH surge lasts 48–50 hours. Because of the careful timing involved in in vitro fertilization programs, we have available some interesting data. 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. 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.

The LH surge initiates the continuation 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, 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. LH-induced cyclic AMP activity overcomes the local inhibitory action of oocyte maturation inhibitor (OMI) and luteinization inhibitor (LI). LI may be endothelin-1, a product of vascular endothelial cells. OMI originates from the granulosa cells, and its activity depends upon an intact cumulus oophorus. Activin also suppresses progesterone production by luteal cells, providing yet another means of preventing premature luteinization.

There is abundant evidence that the oocyte exerts control over granulosa functions. 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.

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. 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. Thus, before and after ovulation, the inhibitor activity is high, while 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 thecal 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. 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.

In the rat, the gene that encodes for plasminogen activator precursor may play a role in various species for several 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.

Prostaglandins of the E and F series and other eicosanoids (especially HETEs, hydroxyeicosatetraenoic acid methyl esters) increase markedly in the preovulatory follicular fluid, reaching a peak concentration at ovulation. Prostaglandin synthesis is stimulated by interleukin-1β, implicating this cytokine in ovulation. Inhibition of the synthesis of these products from arachidonic acid blocks follicle rupture without affecting the other LH-induced processes of luteinization and oocyte maturation. Prostaglandins may act to free proteolytic enzymes within the follicular wall, and the HETEs may promote angiogenesis and hyperemia (an inflammatory-like response). 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 role of prostanoids is so well demonstrated that infertility patients should be advised to avoid the use of drugs that inhibit prostaglandin synthesis. 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. The accumulation of leukocytes is mediated by chemotactic mechanisms of the interleukin system. These immune cells probably contribute to the cellular changes associated with ovulation, corpus luteum function, and apoptosis.

Estradiol levels plunge as LH reaches its peak. This may be a consequence of LH down-regulation of its own receptors on the follicle. Thecal 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 F450c17, 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. 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. Thus, both gonadotropins appear to be involved 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 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. 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. Another possibility has been suggested, a so-called gonadotropin surge-inhibiting factor (GnSIF) originating in the ovary. 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. The administration of a progestin agonist to this experimental model restored 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 Ovulatory Events
1. The LH surge stimulates continuation of reduction division 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, and
4. 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 role of prostanoids is so well demonstrated that infertility patients should be advised to avoid the use of drugs that inhibit prostaglandin synthesis.
plasmin, and to ensure that sufficient LH receptors are present to allow an adequate normal luteal phase.

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. Angiogenesis is an important feature of the luteinization process, a response to LH mediated by growth factors produced in luteinized granulosa cells, such as vascular endothelial growth factor (VEGF). 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 flow units per mass unit 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. 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. 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. 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. There is no evidence that other luteotropic hormones, such as prolactin, play a role in primates during the menstrual cycle. The corpus luteum is not homogeneous, even the large luteal cells, also present are endothelial cells, leukocytes, and fibroblasts. The nonsteroidogenic cells form the bulk (70–85%) of the total cell population. The leukocyte immune cells produce several cytokines, including interleukin-1β and tumor necrosis factor-α. 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 50% of the cells in a mature corpus luteum. 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 now indicated that endothelin-1 may be a mediator of luteolysis.

Even the luteal population is not homogeneous, being composed of at least two distinct cell types, large and small cells. 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. 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, inhibit, and other growth factors) and are more active in steroidogenesis, with greater aromatase activity and more progesterone synthesis than small cells. 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 P450scC and 3β-hydroxysteroid dehydrogenase markedly increases as expected, to account for the increasing production of progesterone, and the continued expression of mRNAs for these enzymes requires LH. 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. Progesterone acts both locally and centrally to suppress new follicular growth. If progesterone concentrations are monitored in ovarian venous effluents following luteectomy in the monkey, ovulation in the subsequent cycle uniformly occurs on the side opposite the higher progesterone level and contralateral to the previous corpus luteum. If circulating progesterone levels are maintained after luteectomy, the subsequent ovulation again occurs in the large cells that have a lower progesterone concentration in its venous effluent. Under normal circumstances (i.e., regular, 28-day cycles), a woman may ovulate from alternate sides.

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. The circulating levels of inhibin A rise in the late follicular phase to reach a peak level at the midluteal phase. 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.

The secretion of progesterone and estradiol during the luteal phase is episodic, and the changes correlate closely with LH pulses. Because of this episodic secretion, relatively low midluteal progesterone levels, which some believe are indicative of an inadequate luteal phase, can be found in the course of totally normal
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 (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; however, the morphological regression of luteal cells may be induced by the estradiol produced by the corpus luteum. There is evidence to support a role for estrogen in the decline of the corpus luteum. 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.

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 (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; however, the morphological regression of luteal cells may be induced by the estradiol produced by the corpus luteum. There is evidence to support a role for estrogen in the decline of the corpus luteum. 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.

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.

Auleta postulated that prostaglandin F\(_2\alpha\) produced within the ovary bearing the corpus luteum or within the corpus luteum serves as the luteolytic agent, and the production of the prostaglandin is initiated by the luteostrogen. Experiments using inhibitors of prostaglandin synthesis have not been helpful because luteal tissue also produces members of the prostaglandin family that have stimulating effects (such as PGE and prostacyclin).

Experimental evidence indicates that the luteolytic effect of prostaglandin F\(_2\alpha\) is mediated by endothelin-1. Prostaglandin F\(_2\alpha\) stimulates the synthesis of endothelin, and endothelin-1 inhibits luteal steroidogenesis. In addition, endothelin-1 stimulates the release of tumor necrosis factor-\(\alpha\), a growth factor known to induce apoptosis.

A newly appreciated characteristic of the corpus luteum has emerged, the importance of 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. 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.

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.

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. Others have indicated that HCG can increase TIMP production, and this, too, would inhibit MMP activity and luteolysis. 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. This new stimulus first appears at the peak of corpus luteum development (9–13 days after ovulation), just in time to prevent luteal regression. HCG serves to maintain the vital steroidogenesis of the corpus luteum until approximately the 9th or 10th week of gestation, by which time placental steroidogenesis is well established. In some pregnancies placental steroidogenesis will be sufficiently established by the 7th week of gestation.

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 3b-hydroxysteroid dehydrogenase) are maximal at...
ovulation and decline throughout the luteal phase. 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. Progesterone acts both centrally and within the ovary to suppress new follicular growth.
3. Regression of the corpus luteum may involve the luteolytic action of its own estrogen production, mediated by an alteration in local prostaglandin and endothelin-1 concentrations.
4. 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. Using a sensitive FSH bioassay, an increase in FSH bioactivity can be measured beginning as early as the midluteal phase. 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.

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. 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. 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. 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.

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. Thus, suppression of FSH secretion during the follicular phase is an action exerted by Inhibin B, whereas escape of FSH inhibition during the luteo-follicular transition is partly a response to decreasing inhibin A secretion by the corpus luteum.

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 if it has an endocrine role. Nevertheless, the timing is right for activin to contribute to the rise in FSH during the luteo-follicular transition. Activin enhances and follistatin suppresses GnRH activity. Evidence in vivo and in vitro indicates that gonadotropin response to GnRH requires activin activity.

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. A progressive and rapid increase in GnRH pulses (as assessed by the measurement of LH pulses) occurs during the luteal-follicular transition. 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. 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 mean 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 inhibition and a concomitant increase in LH pulse frequency. Nevertheless, the timing is right for activin to contribute to the rise in FSH during the luteo-follicular transition. Activin enhances and follistatin suppresses GnRH activity. Evidence in vivo and in vitro indicates that gonadotropin response to GnRH requires activin activity.

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. Our best information comes from two longitudinal studies (with very similar results) of the study of Vollman of more than 30,000 cycles recorded by 650 women and the study of Trenado of more that 25,000 woman-years in a little over 2700 women. The observations of Vollman and Trenado documented a normal evolution in length and variation in menstrual cycles. 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 highest incidence of anovulatory cycles is under age 20 and over age 40. 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. Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.
The duration of the follicular phase is the major determinant of cycle length. Sherman and Korenman predicted in 1975 that a factor other than estrogen is the key — inhibin. 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. 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. Approximately 2–4 years (6–8 years prior to Trenolol) prior to menopause, the cycles lengthen again. In the last 10–15 years before menopause, there is an acceleration of follicular loss. This accelerated loss begins when the total number of follicles reaches approximately 25,000, a number reached in normal women at age 37–38. Eventually menopause occurs because the supply of follicles is depleted.

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). 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.

Chapter References

8. Schipper I, Hop WCJ, Fauser BCJM, The follicle-stimulating hormone (FSH) threshold/window concept examined by different interventions with exogenous FSH during the follicular phase of the normal menstrual cycle: duration, rather than magnitude, of FSH increase affects follicle development, J Clin Endocrinol Metab 83:1295, 1998.


Among his many accomplishments, Galileo Galilei gave to science, in 1609, two important instruments, the telescope and the microscope. Antonj 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 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.

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. The importance of temperature is emphasized by the correlation of reduced numbers of sperm associated with episodes of body fever. 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. Indeed, the injection of one sperm directly into an oocyte (intracytoplasmic sperm injection) with sperm obtained by testicular biopsy (in men with congenital absence of epididymides) is very successful in achieving fertilization and pregnancy. However, the use of sperm from men with sperm abnormalities, and especially the use of immature germ cells for intracytoplasmic injection should be pursued with some caution. 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.

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. 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.
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 mucus microstructure that are smaller than the sperm head; therefore, the sperm must actively push their way through the mucus. 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. 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. It is generally believed that the cervical mucus has a filtering action; abnormal and less "capable" sperm have difficulty getting through. Uterine contractions and sperm motility propel the sperm upward, and in the human, sperm can be found in the tube 5 minutes after insemination. 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. 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. 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.

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. In animals, the fertilizable lifespan is usually one-half the motile lifespan.

The attrition in sperm numbers from vagina to tube is substantial. Of an average of 200 to 300 million sperm deposited in the vagina, at most only a few hundred, and often less, achieve proximity to the egg. 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. Sperm are not stored in the fallopian tube, and indeed, many sperm continue past the oocyte to be lost into the peritoneal cavity. However, the cervix does serve as a reservoir providing a supply of sperm for up to 72 hours.

Within the fallopian tube ampulla, sperm display a new pattern of movement that has been called hyperactivated motility. 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. 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 upon 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, and perhaps completed, during the sperm's passage through the cervix.
The acquisition of hypermotility.

The acrosome reaction is due to the modification and breakdown, followed by a merger, of the sperm cell membrane and the outer acrosomal membrane. Capacitation, a process initiated during the sperm's passage through the cervix or during in vitro incubation in an appropriate medium, is characterized by the following:

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 their surface charge, and restriction of receptor mobility. This is 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. 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. In addition, capacitation endows the sperm with hypermotility, and the increased velocity of the sperm is a very critical factor in achieving zona penetration.

The events that constitute the process of capacitation are regulated by the redox status of the sperm cell. 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.

Although capacitation classically has been defined as a change sperm undergo in the female reproductive tract, 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. The loss of cholesterol regulates the expression of sperm cell membrane surface lectins that are involved in sperm surface receptors for the zona pellucida. The time required for in vitro capacitation is approximately 2 hours.

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 Amplitude

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 or during in vitro incubation in an appropriate medium, is characterized by the following:
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. 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, the cumulus and oocyte are in the ampulla of the fallopian tube.

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.

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.
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. 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 cytokline components during the passage of the egg, perhaps providing important metabolic factors for transport and implantation.

Some have described a menstrual cycle variation in ciliated cells in the tube, but most have not. The apparent variation is probably due to differential changes in response to cyclic hormone levels. 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. 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 moving the egg. However, interference with muscle contractility in the rabbit did not block egg transport. 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. This suggests that ciliary beating is crucial for egg transport. There are fertile 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. However, motility of cilia in the tube may be disordered and not totally absent.

Transvaginal endoscopic observation of actual ovum and cumulus oophorous pickup in a woman 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. 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 E series relax tubal muscle, whereas those of the F series stimulate muscle activity of the tube. Although PGF2α stimulates human oviductal motility in vivo, it does not cause acceleration of ovum transport.

Is there an essential anatomical segment of the tube? Excision of the ampullary-isthmic junction in rabbits does not prevent fertility. 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. The fallopian tube appears to readily adapt to anatomical 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. This implies that there may be a component in uterine fluid during the first 48 hours following ovulation that is toxic to the egg. 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. 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 changing 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. 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
After ovulation, the oocyte and its surrounding cumulus are in the ampulla of the fallopian tube within 2–3 minutes. Tubal transport depends on smooth muscle contractions and ciliary-induced flow of secretory fluid. 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 is regulated by the sex hormones, and 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. 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 spermatozoon. 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 intracellular calcium in overall oocyte fertilization.

### Fertilization

The fertilizable life of the human oocyte is unknown, but most 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. The majority of pregnancies occur when coitus takes place within the 3-day interval just prior to ovulation.

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. This chemotactic responsiveness of sperm requires the changes that take place in the capacitation process. Thus, this may be a system to select a sperm that is fully capable of fertilization.

The cumulus oophorus undergoes a preovulatory expansion that may have 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. It has been suggested, based upon 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. 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 barrier to polyspermy.

Penetration through the zona is rapid and mediated by acrosin, a trypsinlike proteinase that is bound to the inner acrosomal membrane of the sperm. 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, interconnected 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 active motility not only of the tail but of the head as well. Indeed, the sperm head undergoes rapid lateral oscillations of the head about a fulcrum at the head-tail junction, suggesting a scyphelike action on the zona.

The acrosome is a lysosomelike 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. This reaction requires an influx of calcium ions, the eflux of hydrogen ion, 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 within the 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, and ZP3, with ZP3 being the most abundant. ZP3 is the primary ligand for sperm and ZP2 binding occurs after the acrosome reaction, participating in the zona reaction to prevent polyspermy. 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 ZP3 gene is expressed only in growing oocytes. DNA sequence similarities of the ZP3 gene in various mammals indicates that this gene has been evolutionarily conserved and that the sperm-ligand interaction is a common mechanism among mammals. Mice with a disrupted ZP3 gene produce oocytes lacking a zona pellucida and are unable to become pregnant. A vaccine against zona pellucida proteins from pigs is used to control reproduction in female elephants and deer.

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. Once binding is accomplished, the acrosome reaction is triggered by the peptid chain component of the receptor glycoprotein. At least one receptor on the sperm head is a tyrosine kinase, activated by binding to the ZP3 glycoprotein and an initiator of the acrosome reaction. 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...
gestation. In each ovulatory cycle, only 30% of normally fertile couples can achieve a pregnancy. Once conception is achieved, only 30% survive to birth.

Chromosomally abnormal. The fact that only 1 in 200 newborns has a chromosome abnormality attests to the powerful selection mechanisms operating in early human development.

In the postimplantation period, if only clinically diagnosed pregnancies are considered, the generally accepted figure for spontaneous miscarriage in the first trimester is 15%. Approximately 50–60% of these abortuses have chromosome abnormalities. Because the 4 preimplantation losses would not have been recognized clinically, there would have been 6 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. Hendrickx and Kraemer used a similar technique in the baboon and recovered 23 embryos, of which 10 were morphologically abnormal.

Flushing. Two of the 9 embryos were morphologically abnormal and probably would not have implanted. Another example of exocytosis with the cortical reaction, the hardening of the extracellular layer by the zona reaction, is also active in dispelling the cumulus. The cell membrane of the unfertilized oocyte contains integrin adhesion/fusion molecules that recognize peptides such as fibronectin, laminin, and collagen. Fibronectin appears on the zona pellucida of the oocyte. A soluble sperm protein, called oscillin, has been identified in the equatorial segment of the sperm head that may be the signaling agent for the cortical calcium oscillations.

The clinician is 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.

Embryonic genome activity in the human begins early; DNA synthesis activity can be detected 9–10 hours after insemination. Protein kinase C activation is an important step in the acrosomal reaction, leading to phosphorylation of sperm proteins involved in the process.

Protein kinase C activation is an important step in the acrosomal reaction, leading to phosphorylation of sperm proteins involved in the process. As with other lysosomal 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. Thus, the zona block to polyspermy is accomplished.

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. 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. This fusion is mediated by specific proteins. Two membrane proteins from the sperm head have been sequenced; one (PH-20) is involved in binding to the zona pellucida, and the other (PH-30, also called fertilin) is involved in fusion with the oocyte. PH-20, with hyaluronidase activity, is also active in dispelling the cumulus.

The second polar body is released and leaves 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 chromat in 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. The stage is set for the critical calcium oscillations.

An arrest of development in this pre-blastocyst cleavage stage is well recognized. Early embryonic signals may be derived from a store of maternal messenger RNAs, termed the "maternal legacy." An arrest of development in this pre-blastocyst cleavage stage is well recognized. The clinician is 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 9 embryos were morphologically abnormal and probably would not have implanted. Hendrickx and Kraemer used a similar technique in the baboon and recovered 23 embryos, of which 10 were morphologically abnormal. This suggests that, in nonhuman primates, some ovolated 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. 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%. When the loss of fertilized oocytes before implantation is included, approximately 46% of all pregnancies end before the pregnancy is clinically perceived.

In the postimplantation period, if only clinically diagnosed pregnancies are considered, the generally accepted figure for spontaneous miscarriage in the first trimester is 15%. Approximately 50–60% of these abortuses have chromosome abnormalities. 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 30% of normally fertile couples can achieve a pregnancy. Once conception is achieved, only 30% survive to birth.

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 acrosomal 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, 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. 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 30 to 200) 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 6, 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. 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. 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, and days 16–19 of cycles stimulated by exogenous gonadotropins. Endometrial receptivity is heralded by formation of pinopodes, surface epithelial microvilli that exhibit a cystic change, appearing and regressing during the window of receptivity. The pinopodes may serve to absorb fluid from the uterine cavity forcing the blastocyst to be in contact with the endometrial epithelium.

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. EPF prior to implantation is apparently produced by the ovary in response to a signal from the blastocyst. 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.

Blastocysts grown in culture produce and secrete human chorionic gonadotropin (HCG), beginning days 7–8 after fertilization. Messenger RNA for HCG can be found in 6- to 8-cell human embryos. 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. Function of the corpus luteum is crucial during the first 7–9 weeks of pregnancy, and luteectomy early in pregnancy can precipitate abortion. Similarly, early pregnancy loss in primates can be induced by injections of anti-HCG serum. 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. It is not surprising that various growth factors are produced by the early embryo.

In rodents and rabbits, implantation can be interrupted by injection of prostaglandin inhibitors. 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 at implantation sites, similar to any inflammatory response. The blastocysts of mice, rabbits, sheep, and cows produce prostaglandins, and prostaglandin E2 release from human blastocysts and embryos has been demonstrated. The secretory endometrial epithelial cells are also a source of prostaglandin E2 but not prostaglandin F2α, 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. Nevertheless, prostaglandin E2 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. In the rabbit, platelet-activating factor also induces the production of early pregnancy factor (discussed above).

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.

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. It is a process that is limited in both time and space. Implantation begins 2–3 days after the fertilized egg enters the uterus on day 18 or 19 of the cycle. Thus, implantation occurs 5–7 days after fertilization. The implantation site in the human uterus is usually in the upper, posterior wall in the midsagittal plane. Implantation consists of three stages: adhesion, invasion, 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. 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 trophoderm prior to zona hatching). By this time, the blastocyst has differentiated into an inner cell mass (embryo) and trophoderm (placenta), both essential for implantation.

The endometrial production of at least 3 cytokines appears to be involved in implantation. 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. LIF displays the same pattern of expression as CSF-1, and mice with an LIF gene mutation have a failure of blastocyst implantation. Blocking the interleukin-1 receptor in mice also prevents implantation. However, the role of interleukin-1 is less clear because mice that are deficient in the interleukin receptor have normal reproduction. 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. In addition, the blastocyst contains receptors for epidermal growth factor.
integrate the extracellular matrix components, especially laminin and fibronectin, which mediate cell adhesion via the adhesion molecules. 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. Thus implantation starts with adherence due to binding with endometrial integrins, 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. It has been suggested that a lack of integrin expression during the implantation window can be a cause of infertility. The blastocyst also expresses integrins in a time sequence and at a site (outgrowing trophoblast cells) that are appropriate for key activity during implantation.

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. 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 on 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. 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 distinguished from the surface of the epithelial cells by flushing the uterus with physiologic solutions.

Invasion and Placentation

In the second week after ovulation, the placenta is formed. 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.

Three types of interactions between the implanting trophoblast and the uterine epithelium have been described. In the 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 invade itself underneath the epithelium. Last, fusion of trophoblast with individual uterine epithelial cells has been identified by electron microscopy in the rabbit. 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 seem to 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.

Urokinase and proteases, trophoblastic enzymes that convert plasminogen to plasmin, are inhibited by HCG, indicating regulation of this process by the embryo.

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. Additional studies in vitro have shown that cells move away from trophoblast in a process called “contact inhibition.” Trophoblast then spreads to fill the spaces vacated by the cocultured cells. Once the intracellular 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 of trophoblastic tissue during early embryogenesis is regulated by the many growth factors and cytokines produced in both fetal and maternal tissues. Invasion of the early trophoblast requires the expression of integrins, stimulated by insulin-like growth factor-II and inhibited by transforming growth factor-β.

Actively migrating trophoblast cells have a different integrin profile than nonmigrating cells, specifically cell surface receptors that preferentially bind laminin. 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. The role of integrin cell surface receptors is not simply to bind to a
The endometrium is prepared for implantation by the complex activity of cytokines, growth factors, and lipids modulated by the sex hormones, especially estrogen and progesterone. Implantation begins with hatching from the zona pellucida about 1–3 days after the morula entered the uterine cavity.

Chapter 4

The early embryo enters the uterine cavity as an 8-cell morula and becomes a 30 to 200-cell blastocyst before implantation. 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. Plasmin activates 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.

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. The decidual cells at the time of implantation contain 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.

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.

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. Plasmin activates 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.

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. The decidual cells at the time of implantation contain 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.

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.

Histamine may initiate the decidual response. Antihistamines given systemically or directly into the uterus prevent the decidual response in rats. This was disputed when other workers found that systemic antihistamines were not effective in preventing the decidual response. However, there are two different receptors for histamines, H1 and H2. These are not blocked by the same agents, and early experiments demonstrating a lack of effect of antihistamines may have utilized only a block to one receptor. Blockage of both receptors in rats is followed by a decrease in the number of implantation sites. Mast cells in the uterus are a major source of histamine, but it is possible that the embryo can also synthesize histamine.

More recently, 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. 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 (TIMP). 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. 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 cytrophoblasts to differentiate into non-invasive syncytiotrophoblasts. Even human chorionic gonadotropin (hCG) may exert a governing force by inhibiting protease activity.

Key Steps in Implantation

1. The early embryo enters the uterine cavity as an 8-cell morula and becomes a 30 to 200-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).

Chapter References

Steroid Hormones in Pregnancy

**Progesterone**
- In its key location as a way station between mother and fetus, the placenta can utilize 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, as well as pregnenolone, is 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 below).

**Relaxin**
- Relaxin

**Other Prostaglandins**
- Prostaglandins
- Thromboxane and Prostacyclin
- Prostaglandin Biosynthesis
- Prostaglandins and the Fetal Circulation
- Induction of Labor and Fetal Breathing
- Prostaglandin and Fetal Lung Maturation

**Prostaglandins and Prostaglandin Inhibition**
- Prostaglandin Inhibition
- Induction of Labor and Cervical Ripening
- Induced Abortion

Steroid Hormones in Pregnancy

Steroidogenesis in the fetoplacental unit does not follow the conventional mechanisms of hormone production within a single organ. Rather, 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.

**Measurement of Estrogen in Pregnancy**
- Measurement of Estrogen in Pregnancy
- Estrogen

**The Fetal Adrenal Cortex**
- The Fetal Adrenal Cortex

**Placental Sulfatase Deficiency**
- Placental Sulfatase Deficiency

**Protein Hormones of Pregnancy**
- Hypothalamic-like Releasing Hormones
- Human Chorionic Gonadotropin (HCG)
- Human Placental Lactogen (HPL)
- Human Chorionic Adrenocorticotropic Hormone
- Human Chorionic Thyrotropin (HCT)
- Human Chorionic Somatotropin

**Measurement of Estrogen in Pregnancy**
- Measurement of Estrogen in Pregnancy
- Estrogen

**Measurement of Estrogen in Pregnancy**
- Measurement of Estrogen in Pregnancy
- Estrogen

**Measurement of Estrogen in Pregnancy**
- Measurement of Estrogen in Pregnancy
- Estrogen

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 not surprising 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, utero-placental 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 will review the mechanisms by which the fetus establishes influence over important events during pregnancy. The important process of lactation is discussed in Chapter 16.

**Prostaglandins**
- Prostaglandins
- Prostaglandin Biosynthesis
- Thromboxane and Prostacyclin
- Prostaglandins and the Fetal Circulation
- Prostaglandins and Fetal Breathing
- Prostaglandins and Fetal Lung Maturation
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. 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. 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 7th week and 10th 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. At term, progesterone levels range from 100 to 200 ng/mL, and the placenta produces about 250 mg per 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) involving the LDL cell membrane receptors, a process enhanced in pregnancy by estrogen. 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. There is evidence in the baboon that estrogen (estradiol) regulates progesterone production in the placenta. 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.

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. 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 concentration of deoxycorticosterone (DOC) at term is 1200 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. This activity is significant during pregnancy because the rate is proportional to the plasma concentration of progesterone. The fetal kidney is also active in 21-hydroxylation of the progesterone secreted by the placenta into the fetal circulation. At the present time, there is no known physiologic role for DOC during pregnancy.

Progesterone has a role in parturition as will be discussed later in this chapter. It has been suggested that progesterone is important in suppressing the maternal immunological response to fetal antigens, thereby preventing maternal rejection of the trophoblast. 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. 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 about 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) obtained from the fetal circulation and synthesized in the fetal liver. The fetal zone in the adrenal gland is extremely active, but produces steroids with a 3b-hydroxy-D configuration like pregnenolone and dehydroepiandrosterone, rather than 3-keto-D-products such as progesterone. The fetus, therefore, lacks significant activity of the 3b-hydroxysteroid dehydrogenase, D-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. 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. The basic precursors of estrogens are 19-carbon androgens. However, there is a virtual absence of 17α-hydroxylase 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. 11

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 (DHAS) production by the fetal adrenal gland. 20 The high output of DHAS by the fetal zone is due to low 3b-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 will be discussed in greater detail later in this chapter.

The fetal adrenal provides DHAS as precursor for placental production of estrone and estradiol. However, the placenta lacks 16α-hydroxylation ability, and estradiol 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 sulfate for placental estradiol formation. After birth, neonatal hepatic 16α-hydroxylation activity rapidly disappears. The maternofetal contribution of DHAS 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 DHAS 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. 24

The profiles of the unconjugated compounds in the maternal compartment for the 3 major estrogens in pregnancy are:

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. 25,26 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 DHAS and fetal DHAS, and its

![Graph of Maternal Plasma Unconjugated Estrogens over Pregnancy Weeks](image)
During pregnancy, estrone and estradiol excretion is increased about 100 times over nonpregnant levels. However, the increase in maternal estradiol excretion is about a thousand-fold. The traditional view that estradiol 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. Because of its high production rate and concentration, estradiol is an important hormone in pregnancy. The maternal level of estradiol is higher than in the fetus; in contrast, the estradiol 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. Blood volume is increased by estrogen stimulation of the maternal and trophoblastic renin-angiotensin systems, and uteroplacental blood flow, that 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. 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. 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. 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 6–10% of the maternal blood estriol is unconjugated.

The Fetal Adrenal Cortex

The fetal adrenal cortex is differentiated by 8–9 weeks gestational age into a thick inner fetal zone and a thin outer definitive zone, the source of cortisol and the forerunner of the adult cortex. 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 80% of the bulk of the gland) rapidly involutes to be replaced by simultaneous expansion of the adrenal cortex composed of the zona glomerulosa, the zona fasciculata, and the zona reticularis. 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 (DHA) and DHAS production rises steadily concomitant with the increase in the size of the fetal zone and adrenal weight. DHA and DHAS are the major secretory products of the fetal zone because 3b-hydroxysteroid dehydrogenase-isomerase activity and the expression of this enzyme's gene are suppressed. The well-known increase in maternal estrogen levels is significantly influenced by the increased availability of fetal DHAS as a precursor. Indeed, the accelerated rise in maternal estrogen levels near term can be explained, in part, by an increase in fetal DHAS. The stimulus for the substantial adrenal growth and steroid production has been a puzzle.

Activin contains two subunits that are identical to the beta-subunits of the inhibins. Inhibin consists of two dissimilar peptides, alpha- and beta-subunits, linked by disulfide bonds. Inhibin A and inhibin B each contain an identical alpha-subunit, but only in inhibin B is a distinct beta-subunit present. The CYP19 gene is also expressed in the fetal liver. The placenta prevents maternal cortisol from reaching the fetus by converting cortisol to cortisone. This 11b-hydroxysteroid dehydrogenase activity is stimulated by placental estrogen.

The Fetal Adrenal Cortex

The fetal adrenal cortex is differentiated by 8–9 weeks gestational age into a thick inner fetal zone and a thin outer definitive zone, the source of cortisol and the forerunner of the adult cortex. 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 80% of the bulk of the gland) rapidly involutes to be replaced by simultaneous expansion of the adrenal cortex composed of the zona glomerulosa, the zona fasciculata, and the zona reticularis. 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 (DHA) and DHAS production rises steadily concomitant with the increase in the size of the fetal zone and adrenal weight. DHA and DHAS are the major secretory products of the fetal zone because 3b-hydroxysteroid dehydrogenase-isomerase activity and the expression of this enzyme's gene are suppressed. The well-known increase in maternal estrogen levels is significantly influenced by the increased availability of fetal DHAS as a precursor. Indeed, the accelerated rise in maternal estrogen levels near term can be explained, in part, by an increase in fetal DHAS. The stimulus for the substantial adrenal growth and steroid production has been a puzzle.

Inhibin-A: Alpha-Beta_A
Inhibin-B: Alpha-Beta_B

The Three Forms of Activin (Homodimers)

Activin-A: Beta_A-Beta_A
Activin-AB: Beta_A-Beta_B
Activin-B: Beta_B-Beta_B

The Two Forms of Inhibin (Heterodimers)

The CYP19 gene is expressed in the fetal liver. The placenta prevents maternal cortisol from reaching the fetus by converting cortisol to cortisone. This 11b-hydroxysteroid dehydrogenase activity is stimulated by placental estrogen. Regulation of this enzyme by estrogen influences fetal ACTH secretion. With internalization of LDL-cholesterol, hydrolysis by lysosomal enzymes of the cholesterol 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. Thus, near term, both de novo synthesis and utilization of LDL-cholesterol are necessary to sustain the high rates of DHAS and estrogen formation. In addition, ACTH increases adrenal response by increasing the expression of its own receptor.

The tropic support of the fetal adrenal gland by ACTH from the fetal pituitary is protected by placental estrogen. The placenta prevents maternal cortisol from reaching the fetus by converting cortisol to cortisone. This 11b-hydroxysteroid dehydrogenase activity is stimulated by placental estrogen. Regulation of this enzyme by estrogen influences fetal ACTH secretion. With internalization of LDL-cholesterol, hydrolysis by lysosomal enzymes of the cholesterol 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. Thus, near term, both de novo synthesis and utilization of LDL-cholesterol are necessary to sustain the high rates of DHAS and estrogen formation. In addition, ACTH increases adrenal response by increasing the expression of its own receptor.

The Three Forms of Activin (Homodimers)

Activin-A: Beta_A-Beta_A
Activin-AB: Beta_A-Beta_B
Activin-B: Beta_B-Beta_B

The Two Forms of Inhibin (Heterodimers)

Inhibin-A: Alpha-Beta_A
Inhibin-B: Alpha-Beta_B

It has been suggested that rising fetal cortisol secretion 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. 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 DHAS to serve as precursor for the increase in estrogen that occurs prior to parturition. Another possibility is CRH direct stimulation of DHAS by the fetal adrenal gland, as has been demonstrated in vitro.

Adrenal gland steroidogenesis involves autoinhibitory and paracrine regulation by various growth factors. Fetal adrenal cells produce inhibin, and the a-subunit (present only in inhibin) is preferentially decreased by ACTH. In the fetal adrenal, the beta-subunit is not expressed; thus, inhibin A and activin A are the principal forms. Inhibin consists of two dissimilar peptides, alpha- and beta-subunits, linked by disulfide bonds. Inhibin A and inhibin B each contain an identical alpha-subunit, but distinct beta-subunits. Activin contains two subunits that are identical to the beta-subunits of the inhibins.

The Two Forms of Inhibin (Heterodimers)

Inhibin-A: Alpha-Beta_A
Inhibin-B: Alpha-Beta_B

The Three Forms of Activin (Homodimers)

Activin-A: Beta_A-Beta_A
Activin-AB: Beta_A-Beta_B
Activin-B: Beta_B-Beta_B

It has been suggested that rising fetal cortisol secretion 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. 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 DHAS to serve as precursor for the increase in estrogen that occurs prior to parturition. Another possibility is CRH direct stimulation of DHAS by the fetal adrenal gland, as has been demonstrated in vitro.

Adrenal gland steroidogenesis involves autoinhibitory and paracrine regulation by various growth factors. Fetal adrenal cells produce inhibin, and the a-subunit (present only in inhibin) is preferentially decreased by ACTH. In the fetal adrenal, the beta-subunit is not expressed; thus, inhibin A and activin A are the principal forms. Inhibin consists of two dissimilar peptides, alpha- and beta-subunits, linked by disulfide bonds. Inhibin A and inhibin B each contain an identical alpha-subunit, but distinct beta-subunits. Activin contains two subunits that are identical to the beta-subunits of the inhibins.
Amniotic Fluid Estrogen Measurements utilizes 3 markers in the maternal circulation: α-fetoprotein, human chorionic gonadotropin, and unconjugated estriol. Radioimmunoassay of unconjugated estriol in the plasma. 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 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. Measurement of maternal estrogen levels. To any of the key organs in the fetus, placenta, and mother becomes important. 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. Focus 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 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 estradiol because 90% of maternal estradiol 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 as well as 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. Fetal hypoxemia due to reduced uteroplacental blood flow is associated with a reduction in fetal androgen production. Treatment of the neonatal baboon with estradiol did not halt the regression of the fetal zone and DHAS production, arguing against the hypothesis that the fetal zone is dependent on an estrogen-induced deficiency in 3β-hydroxysteroid dehydrogenase. Treatment of the pregnant baboon with estradiol reduced production of DHAS. It continues to be uncertain, however, whether the internal microenvironment of the adrenal gland can be affected by the exogenous administration of steroids. In the monkey, epidermal growth factor can increase the 3β-hydroxysteroid dehydrogenase content in the fetal adrenal gland, but it is not clear how this action is regulated. This explanation (estrogen regulation of 3β-hydroxysteroid dehydrogenase) is further challenged by in vitro studies of human fetal zone cells indicating that estradiol and IFG-II combine to direct steroidogenesis to DHAS in a mechanism not due to inhibition of 3β-hydroxysteroid dehydrogenase. Nevertheless, it is an attractive and useful hypothesis to view the principal mission of the fetal adrenal as providing DHAS as the basic precursor for placental production. Estrogen, in turn, feeds back to the adrenal to direct steroidogenesis along the D pathway to provide even more of its precursor, DHAS. Thus far, this is the only known function for DHAS. With birth and loss of exposure to estrogen, the fetal adrenal gland quickly changes to the adult type of gland. 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 estradiol because 90% of maternal estradiol 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 as well as 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. Fetal hypoxemia due to reduced 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. 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 radioimmunoassay of unconjugated estradiol in the plasma. Because of its short half-life (5–10 minutes) in the maternal circulation, unconjugated estradiol has less variation than urinary or total blood estradiol. However, assessment of maternal estradiol levels has been superseded by various biological 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 3 markers in the maternal circulation: alpha fetoprotein, human chorionic gonadotropin, and unconjugated estradiol. Amniotic Fluid Estrogen Measurements Amniotic fluid estradiol is correlated with the fetal estrogen pattern rather than the maternal. Most of the estradiol in the amniotic fluid is present as 16-glucosiduronate or...
as 3-sulfate-16-glucosuronate. 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 glucosuronate conjugates are removed slowly from the fetus. The glucosurononates therefore predominate in the fetal urine and the amniotic fluid. Because of the slow changes in glucosurononates, 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.

**Estriol**

Estriol (15a-hydroxyestradiol) is formed from a fetal precursor, and is very dependent upon 15a-hydroxylation activity in the fetal liver. The capacity for 15a-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. There is no clinical use for maternal blood or urine estriol measurements during pregnancy. The clinical use of maternal blood and urine estriol measurements is of no advantage over the usual estriol assessment.

**Placental Sulfatase Deficiency**

There is an X-linked metabolic disease characterized by a placental sulfatase deficiency in the syncytiotrophoblast, and postnatally, ichthyosis, occurring in about 1 in 2000–3000 newborn males. Patients with the placental sulfatase disorder are unable to hydrolyze DHAS or 16a-hydroxy-DHAS, and, 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 dilatation; 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 has been 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. 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). 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. 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 DHAS and normal amniotic fluid DHA and androstenedione. The normal DHA and androstenedione with a high DHAS rule out congenital adrenal hyperplasia. The small amount of estriol that is present in these patients probably arises from 16a-hydroxylation of DHAS in the maternal liver, thus providing 16a-hydroxylated DHA 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 DHAS in the amniotic fluid is reliable. 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:cholesterol ester ratio (due to the accumulation of cholesterol sulfate). This inherited disorder, thus, represents a single entity: placental sulfatase deficiency and X-linked ichthyosis, occurring in about 1 in 2000–3000 newborn males.

There is an X-linked metabolic disease characterized by a placental sulfatase deficiency in the syncytiotrophoblast, and postnatally, ichthyosis, occurring in about 1 in 2000–3000 newborn males. Patients with the placental sulfatase disorder are unable to hydrolyze DHAS or 16a-hydroxy-DHAS, and, 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 dilatation; 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 has been 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. 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). 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. 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 DHAS and normal amniotic fluid DHA and androstenedione. The normal DHA and androstenedione with a high DHAS rule out congenital adrenal hyperplasia. The small amount of estriol that is present in these patients probably arises from 16a-hydroxylation of DHAS in the maternal liver, thus providing 16a-hydroxylated DHA 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 DHAS in the amniotic fluid is reliable. 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:cholesterol ester ratio (due to the accumulation of cholesterol sulfate). This inherited disorder, thus, represents a single entity: placental sulfatase deficiency and X-linked ichthyosis, occurring in about 1 in 2000–3000 newborn males.

**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 syncytiotrophoblast 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. The protein hormone system is complicated because individual peptides can have multiple functions. 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**

<table>
<thead>
<tr>
<th>Fetal Compartment</th>
<th>Placental Compartment</th>
<th>Maternal Compartment</th>
</tr>
</thead>
</table>

- [Image for visual aids]
Hypothalamic-like hormones

- GnRH
- CRH
- TRH
- Somatostatin
- GHRH

Pituitary-like hormones

- HCG
- HPL
- HGH
- HCT
- ACTH
- Oxytocin

Growth factors

- IGF-I
- IGF-II
- Epidermal growth factor
- Platelet-derived growth factor
- Transforming growth factor-a
- Transforming growth factor-b
- Activin
- Follistatin

Cytokines

- Interleukins
- Interferons
- Tissue necrosis factor-a
- Colony stimulating factor-1

Other

- Opiates
- Prorenin
- Pregnancy-specific b-glycoprotein
- Pregnancy-associated plasma protein A

<table>
<thead>
<tr>
<th>Alpha-fetoprotein</th>
<th>Hypothalamic-like hormones</th>
<th>Decidual proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GnRH</td>
<td>Prolactin</td>
</tr>
<tr>
<td></td>
<td>CRH</td>
<td>Relaxin</td>
</tr>
<tr>
<td></td>
<td>TRH</td>
<td>Interleukin-1</td>
</tr>
<tr>
<td></td>
<td>Somatostatin</td>
<td>Colony stimulating factor-1</td>
</tr>
<tr>
<td></td>
<td>GHRH</td>
<td>Progesterone-associated endometrial protein</td>
</tr>
</tbody>
</table>

HCG: Human Chorionic Gonadotropin

The human placenta contains many releasing and inhibiting hormones, including gonadotropin-releasing hormone (GnRH), corticotropin-releasing hormone (CRH), thyrotropin-releasing hormone (TRH) and somatostatin. 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. 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. The placental receptors for GnRH have a lower affinity than that of GnRH receptors in the pituitary, ovary, and testis. 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. GnRH release is increased by estrogen, activin A, insulin, and prostaglandins, and inhibited by progesterone, endogenous opiates, inhibin, and follistatin.

CRH, identical in structure to hypothalamic CRH, is produced in the trophoblast, the fetal membranes, and the decidua. Its production is regulated by steroids, decreased by progesterone and, in contrast to the usual negative feedback action in the hypothalamus, increased by glucocorticoids. These interactions are consistent with the increase in fetal and maternal ACTH and cortisol associated with the last weeks of pregnancy and 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. 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. Maternal CRH levels are elevated in women with pregnancies under stress; e.g., with preeclampsia and preterm labor. The increase in placental CRH may be a response to the activation of fetal pituitary ACTH and adrenal cortisol secretion in the presence of hypoxemia.

Trophoblast, amnion, chorion, and decidua also produce a peptide similar to CRH, named urocortin, that binds to CRH receptors and CRH-binding protein.

Human Chorionic Gonadotropin (HCG)

Human chorionic gonadotropin is a glycoprotein, a peptide framework to which carbohydrate side chains are attached. 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 (a) and beta (b). The a-subunit in these glycoprotein hormones is identical, consisting of 92 amino acids. Unique biological activity as well as specificity in radioimmunoassays is attributed to the molecular and carbohydrate differences in the b-subunits (see "Heterogeneity" in Chapter 2).

HCG is the largest b-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 b-HCG contains 4 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.
HCG gene expression is present in both cytotrophoblast and syncytiotrophoblast, but it is synthesized mainly in the syncytiotrophoblast. 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.

These differences in structure are associated with a different promoter and transcriptional site that is located upstream in the HCG b-subunit gene compared with the transcriptional site in the LH b-subunit gene. The HCG b-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.

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. Using recombinant DNA technology, it has been demonstrated that there is a single human gene for the expression of the a-subunit. The gene for the a-subunit is 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 a-gene in both placenta and pituitary. The a-subunit gene is expressed in several different cell types, but the b-subunit genes are restricted in cell type. The TSH b-gene is expressed only in thyrotrophs regulated by thyroid hormone; the FSH b-gene is expressed in gonadotrophs regulated by GnRH, activin, and gonadal steroids; the LH b-gene, also expressed in gonadotrophs, is regulated by GnRH and is unaffected by activin and inhibit.

The a-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 a-gene expression. In gonadotrophs, the GnRH-signaling pathway for a-gene transcription utilizes phosphorylase-stimulation of dacyl glycerol (DAG) and inostol trisophosphate (IP3) that lead to a release of intracellular calcium stores. GnRH also stimulates the influx of calcium at the cell membrane. DAG, IP3, and calcium work together to stimulate protein kinase C activity. Protein kinase regulation of the a-promoter is a principal part of the overall mechanism. This 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 b-subunits of LH, HCG, and TSH are located in a cluster on chromosome 19q13.3. There are 6 genes for the b-subunit of HCG, and only one for LH. Transcription for the 6 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 b-HCG evolved relatively recently from b-LH, and the unique amino acid terminal extension of b-HCG arose by a read-through mutation of the translation stop codon in the b-LH gene; the DNA sequences of the b-HCG genes and the b-LH gene are 96% identical. Only primates and horses have been demonstrated to have genes for the b-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 b-subunit that are homologous to those in the pituitary gonadotropins of other species. The equine b-chorionic gonadotropin gene is identical to the equine b-LH gene, and although the primate b-HCG gene evolved from the same ancestral b-LH gene, the horse chorionic gonadotropin gene evolved in a different way. The b-LH gene is not expressed in the placenta.

The placenta-specific expression of b-HCG is due to several differences in DNA sequences between the b-HCG and b-LH genes. The cyclic AMP-mediated enhancement of the b-HCG promoter is influenced by several regulatory proteins. The study of the b-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 b-HCG genes.

The genetic complexity for the transcription of b-HCG raises the possibility for mutations of these genes as causes of reproductive problems. A search for b-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. 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. Similar responses can be demonstrated with other peptides, such as interleukin-1b. Similar to opiate action in the hypothalamus, the endorphins are a major inhibiting influence on HCG secretion. 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 together by progesterone. Follistatin, by binding activin, prevents the stimulatory activity of activin. Other growth factors, specifically IGFI, IGF-II, TGF-a, 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. Inhibin inhibition in the placenta is further stimulated by prostaglandins.

Can the cytotrophoblast-syncytiotrophoblast relationship be compared with the hypothalamus-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 8th day after ovulation, one day after implantation, when b-HCG first can be detected in maternal blood. HCG has been detected at the 8-cell stage in the embryo using molecular biology techniques. Continued survival of the corpus luteum is totally dependent on HCG, and, in turn, survival of the pregnancy is dependent upon steroids from the corpus luteum until the 7th week of pregnancy. From the 7th week to the 10th week, the corpus luteum is gradually replaced by the placenta, and by the 10th 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. 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 b-HCG gene is expressed in fetal kidney and fetal adrenal, suggesting that HCG may affect the development and function of these organs.

HCG gene expression is present in both cytotrophoblast and syncytiotrophoblast, but it is synthesized mainly in the syncytiotrophoblast. The maternal circulating
When the HCG titer exceeds 1000–1500 IU/L, vaginal ultrasonography should identify the presence of an intrauterine gestation. The quantitative measurement of HCG can assess pregnancy viability. A normal rate of rise usually indicates a normal pregnancy. 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. A major route of clearance for HCG is renal metabolism in which a final reduced fragment of the b-subunit is produced, known as the b-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 10th week of gestation.

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.

There are two clinical conditions in which blood HCG titers are very helpful: trophoblastic disease and ectopic pregnancies. Early pregnancy is characterized by the sequential appearance of HCG, followed by b-HCG and then a-HCG. The ratio of b-HCG to whole HCG remains constant after early pregnancy. Trophoblastic disease is distinguished by very high b-HCG levels (3–100 times higher than normal pregnancy). Ectopic production of alpha- and beta-HCG by nontrophoblastic tumors is rare. Previous studies with polyclonal antisera suggesting ectopic production were not accurate. The production of whole HCG in such tumors may not occur.

In the United States, hydatidiform moles occur in approximately 1 in 600 induced abortions and 1 in 1000–2000 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. A diagnosis of gestational trophoblastic disease is made when the b-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. In order 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.

In order to avoid unnecessary treatment (prophylactic chemotherapy) of the 80–85% of patients who undergo spontaneous remission, there is a need to identify those at high risk for persistent trophoblastic disease. A radioimmunoassay for the free beta-subunit of HCG may serve this need in that persistent trophoblastic disease is associated with excessive production of the free beta-subunit.

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 32. 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 usually indicates a normal pregnancy.
2. When the HCG titer exceeds 1000–1500 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, and apparently the source of this circulating HCG is the pituitary gland. The concentration of this pituitary HCG normally reaches the sensitivity of the usual modern assay only in a rare postmenopausal woman with high LH levels. HCG produced in sites other than the placenta has little or no carbohydrate, and, therefore, it has a very short half-life and is rapidly cleared from the circulation. Significant levels of free alpha-subunit are also present in the circulation of healthy individuals; however, the levels of the beta-subunit are extremely low.

**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 two 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 and 3 for HPL; however, only 2 of the HPL genes are abundantly active in the placenta, each producing the same HPL hormone. The third HPL gene does generate a protein in the placenta, but its activity is limited.

Although HPL has about 50% of the lactogenic activity of sheep prolactin in certain bioassays, its lactogenic contribution in human pregnancy is uncertain. Its half-life 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 (5–7 mg/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 mg/mL have been found with quadruplets and quintuplets. An abnormally low level is anything less than 4 mg/mL in the last trimester.

**Physiologic Function**

Although 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.
In the mother, HPL stimulates insulin secretion and IGF-I production and induces insulin resistance and carbohydrate intolerance. 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 information and studies in fasted pregnant women have led to the following formulation for the physiologic function of HPL. 

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. 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 (peripheral insulin resistance and hyperinsulinemia).

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.

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.  

The lipid, lipoprotein, and apolipoprotein changes during pregnancy are positively correlated with changes in estradiol, progesterone, and HPL. 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. Fetal undernutrition lowers fetal IGF-I levels, and this is associated with a high prevalence of insulin resistance later as adults. 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. Although some studies indicated that HPL was valuable in screening patients for potential fetal complications, others did not support the use of HPL measurements. Even though utilization 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.

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.
Human Chorionic Thyrotropin (HCT)

The human placenta contains two thyrotropic substances. One is called human chorionic thyrotropin (HCT), similar in size and action to pituitary TSH. Te 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 a-subunit. Antiserum generated to a-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. Studies have indicated that HCG has intrinsic thyrotropic activity, suggesting that HCG is the second placentally thyrotropic substance. It has been calculated that HCG contains approximately 1/4000th of the thyrotropic activity of human TSH. In conditions with very elevated HCG levels, the thyrotropic activity of HCT, and this can be a diagnostic problem in normal pregnancy. There is a correlation between elevated thyroid function and hyperemesis gravidarum. 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). Specifically, HCG with reduced sialic acid content is increased in pregnant patients with hyperemesis and hyperthyroidism. The thyroid hormone changes in pregnancy and the role of HCG as a thyroid stimulator are discussed in Chapter 25.

Human Chorionic Adrenocorticotropin

The rise in maternal free cortisol that takes place throughout pregnancy is due to ACTH and corticotropin-releasing hormone (CRH) production ad secretion into the maternal circulation by the placenta. The placental content of ACTH is higher than can be accounted for by the contribution of sequestrated blood. In addition, cortisol levels in pregnant women are resistant to dexamethasone suppression, suggesting that there is a component of maternal ACTH and CRH that does not originate in the maternal 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. Placental pro-opiomelanocortin (POMC) gene expression and ACTH content are present throughout pregnancy and increase in the weeks before term. One can speculate that placental ACTH and CRH raise maternal adrenal activity in order to provide the basic building blocks (glucocorticoids and pregnenolone) for placental steroidogenesis. There is no passage of ACTH between the fetal and maternal compartments.

The maternal ACTH response to the administration of CRH during pregnancy is blunted, indicating 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. This is further evidence that placentally CRH produces a state of chronic stimulation for the maternal pituitary. In contrast to maternal ACTH, maternal plasma CRH levels are higher in young women, CRH increases in the second trimester to peak values at term. 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. The fall in CRH-binding protein near term further increases the cortisol availability during labor and delivery.

Both maternal and fetal levels of CRH are further elevated in pathologic states such as premature labor, hypertension, fetal asphyxia, and intrauterine growth retardation. Because CRH also stimulates prostaglandin synthesis in the placenta and fetal membranes, it is implicated in the premature labor that accompanies pathologic conditions.

Growth Hormone, Growth Hormone-Releasing Hormone, and Somatostatin

Growth hormone-releasing hormone (GHRH) and somatostatin are found in the placenta, and somatostatin is present in decidua. The amount of somatostatin decreases during gestation. Pregnancy GHRH and somatostatin do not contribute to maternal circulating levels of these peptides. One of the two growth hormone genes on chromosome 17 is expressed only in the syncytiotrophoblast of the placenta. The placental growth hormone is not identical to pituitary growth hormone, and after 15–20 weeks of pregnancy, placentally growth hormone gradually replaces pituitary growth hormone in the maternal circulation. 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. 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 GHRH, but responds inversely to maternal glucose levels, protecting glucose availability for the fetus. Placental growth hormone can also stimulate glucoseogenesis and lipolysis in maternal organs. It is believed, therefore, that placental growth hormone influences fetal growth by affecting maternal metabolism.

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. 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.

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 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, preclampsia, neonatal death, and low birth weight (probably reflecting an increase in villous surface area in response to an adverse intrauterine environment). 34 35

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 b-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, HCG, or b-HCG, and unconjugated estriol. 36 37 This protocol will detect 85% of open neural tube defects and 80% of Down syndrome. However, Down syndrome represents only about 50% of the chromosomal abnormalities that can be detected.

The multiple marker screening protocol measures AFP, HCG, and unconjugated estriol 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 hydroptic fetal Turner syndrome. A comparison trial concluded that a combination of AFP and the free b-subunit of HCG is more accurate and less expensive than the utilization of the 3 markers, AFP, HCG, and unconjugated estriol. 38

The most critical factor for correct risk assessment is accurate gestational dating. A two-week error in dating can change the calculated risk for Down syndrome ten-fold. Therefore, ultrasonographic confirmation of gestational dating is essential. In addition, ultrasonography will indicate fetal number and assess the fetus for anomalies.

The multiple marker protocol is for screening a low-risk population regardless of age, and amniocentesis is necessary for final diagnosis. Genetic amniocentesis is still recommended for older women; however, although multiple marker screening does not detect all chromosomal abnormalities, some have argued that the detection rate is so high that screening as an option should be offered even to older women. 39 40

The detection of a very low unconjugated estriol level in the second trimester is associated with an increased risk of early fetal demise. 34 In addition, a low estriol can
be a marker for placental sulfatase deficiency.

Relaxin

Relaxin is a peptide hormone produced by the corpus luteum of pregnancy, which is not detected in men or nonpregnant women. It is composed of two short peptide chains (24 and 29 amino acids, respectively) linked by disulfide bridges. While 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. The maternal serum concentration rises during the first trimester when the corpus luteum is dominant and declines in the second trimester. 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. 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. To examine the contribution of the corpus luteum, normally pregnant women were compared with women pregnant with donated oocytes (and therefore without corpora lutea). 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 is being tested for ripening of the cervix.

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 electrolytes in the amniotic fluid, and relaxin, which may influence progestagen 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. 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 decidualization proteins (perhaps prolactin itself). 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.

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. 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.

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. 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. There is marked variability in maternal prolactin levels in pregnancy, with a diurnal variation similar to that found in nonpregnant persons.

Amniotic fluid concentrations of prolactin parallel maternal serum concentrations until the 10th 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 minimal maternal and fetal blood levels, yet there is normal fetal growth and development, and amniotic fluid levels are unchanged. 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: choricontal cytotrophoblast, decidua, amnion, and syncytiotrophoblast.

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. Interleukin-1b 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-1b has been localized to mesenchymal fibroblasts from the core of placental villi. Thus, a system of communication is present between maternal decidua and fetal tissue to provide growth factor support for the placenta which 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. Thus, the interleukin-1 influence on HCG secretion is mediated by the interleukin-6 system. The cytokines are derived from interleukin-1 and tumor necrosis factor-a (TNF-a) synergistically release interleukin-6 and activate the interleukin-6 system to secrete HCG. 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. 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 6 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. 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). 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.

How these changes interact to regulate growth and development of different organs and tissues is a new area of research activity. For example, 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. 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. 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. 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. Undoubtedly, the high levels of insulin 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 PGE2 and PGF2α. These effects are dependent on the cytokine milieu and developmental stage. The trophoblast 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. 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. 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 PGE2 and PGF2α. 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. It is not certain whether maternal blood levels of endogenous opiates increase with advancing gestation. However, a marked increase in maternal values is reached during labor, coinciding with full cervical dilatation. 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 circulating levels of renin, the inactive precursor of renin, increase (10-fold) during early pregnancy, the result of ovarian stimulation by HCG. This increase in renin from the ovary is not associated with any significant change in the blood levels of the active form, renin. Possible roles for this ovarian preprorenin-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 tissue and is highly concentrated in the amniotic fluid. The highest biologic levels of preprorenin are found in gestational sacs in early pregnancy; its possible roles in embryonic growth and development remain speculative. Renin and angiotensinogen (the renin substrate) are expressed by the following fetal tissues: chorion, amnion, and placenta. This system responds to a variety of factors, affecting vascular resistance and blood volume. Maternal renin activity is increased four-fold by midgestation, partly a response to an estrogen-induced increase in angiotensinogen, but largely a compensatory response to maintain blood pressure in the presence of vasodilatation. There is little evidence that fetal or uterine prorenin or renin contribute to the maternal circulation.

Attrial Natriuretic Peptide

Atrial natriuretic peptide (ANP) is derived from human atrial tissue and the placenta. ANP 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 maternal side suggest that ANP is a circulating hormone in the fetus. 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.

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 coltus results in a pregnancy. 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. We are just beginning to learn about this fetal and maternal protein. 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.
Pregnancy-specific \textit{g}-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 closely related glycoproteins encoded by genes on chromosome 19. The PSG family is closely related to the carcinoembryonic antigen (CEA) proteins. Pregnancy-associated plasma protein-A (PAPP-A) is a placent al protein that is similar to a macroglobulin in the serum and still in search of specific functions. 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.

**Prostaglandins**

**Prostaglandin Biosynthesis**

The family of prostaglandins with the greatest biologic activity is that having two double bonds, derived from arachidonic acid. 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 lipoxigenase pathway or the cyclooxygenase (prostaglandin endoperoxide H synthase) pathway, depending upon the local cellular context. There are 3 lipooxygenase enzymes that lead to active compounds. 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$_4$. 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 100–1000 times more potent than histamine in the pulmonary airway. Asthma can now be treated with a specific leukotriene receptor antagonist.

![Diagram of Prostaglandin Biosynthesis](image)

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.

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 E$_2$ series, and pentanoic acid, the PG G$_2$ 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$_2$ and possibly PGD$_2$. The a in PGF$_2$ indicates the a steric configuration of the hydroxy 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. 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, and endotoxins, and, therefore, it is called the inducible form. Thus selective inhibition of COX-2 would be therapeutically advantageous, possibly avoiding the side effects associated with inhibition of COX-1.
Thromboxanes and Prostacyclin

Thromboxanes are not true prostaglandins because of the absence of the pentane ring, but prostacyclin (PGI₂) is a legitimate prostaglandin. Thromboxane (TX) (half-life about 30 seconds) and PGI₂ (half-life about 2–3 minutes) can be viewed as opponents, each having powerful biologic activity that counters or balances the other. TXA₂ is the most powerful vasoconstrictor known, whereas PGI₂ is a potent vasodilator. These two agents also have opposing effects on platelet function. Platelets, lungs, and the spleen predominantly synthesize TXA₂ while the heart, stomach, and blood vessels throughout the body synthesize PGI₂. The lungs are a major source of prostacyclin. Normal pulmonary endothelium makes prostacyclin while TXA₂ appears in response to pathologic stimuli. The pulmonary release of prostacyclin may contribute to the body's defense against platelet aggregation.

Let's 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 TXA₂, 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 PGI₂ and its constant presence inhibits platelet aggregation and adherence, keeping blood vessels free of platelets 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 PGI₂ 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 vasodilatation 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.

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 PGI₂. LDL from men and postmenopausal women inhibits and HDL stimulates prostacyclin production. Platelets from diabetics and from class A diabetic pregnant women make more TXA₂ than platelets from normal pregnant women. Smokers who use oral contraceptives have increased platelet aggregation and an inhibition of prostacyclin formation. Incidentally, onion and garlic inhibit platelet aggregation and TXA₂ synthesis. Perhaps the perfect contraceptive pill is a combination of prostogen, 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. The diet of Eskimos and Japanese has a high content of pentanionic acid and low levels of linoleic and arachidonic acids. Pentanionic acid is the precursor of prostaglandin products with 3 double bonds, and, as it happens, PGI₂ is an active agent while TXA₂ is either not formed, or it is inactive. The fat content of most common fish is 8–12% pentanionic 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-PGF₁α, the metabolite of prostacyclin, and TXB₂, the metabolite of thromboxane A₂.

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. 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 lipoxigenase and cyclooxygenase pathways, they are very effective anti-inflammatory agents and antihypersensitivity agents, especially for the treatment of asthma.

Aspirin is an irreversible inhibitor, selectively acetyling the cyclooxygenase involved in prostaglandin synthesis. The other inhibiting agents, nonsteroidal anti-inflammatory agents 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 has no anti-inflammatory properties nor does it affect platelets. However, acetaminophen does reduce prostacyclin synthesis; the reason for this preferential effect is unknown. 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. Diclofenac, acetaminophen, and naproxen inhibit both enzymes equally. The side effects associated with each agent are a reflection of the degree of selectivity towards the two enzymes; inhibition of COX-1, the constitutive form, being associated with significant side effects, and inhibition of COX-2, the inducible form, being therapeutic. 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 PGE₂ protects the gastric mucosa by inhibiting gastric acid secretion, and COX-1 is the predominant enzyme in the gastric mucosa.
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). Prostaglandin 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 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. Others indicate that the dose which effectively and selectively inhibits platelet cyclooxygenase is 20–40 mg daily. 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 TXA2 synthesis would be superior to aspirin in terms of antithrombotic effects.

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.

Extensive work in the sheep has implicated the fetal pituitary-adrenal axis in normal parturition. The sequence of events in the sheep 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 sheep with a decline in progesterone. This change is brought about by the induction of 17a-hydroxylase, 17,20-lyase enzyme activity (P450c17) in the sheep placenta. The up-regulation of P450c17 may be mediated by PGE2. COX-2 activity is stimulated by cortisol, while at the same time, cortisol inhibits the activity of 15-hydroxyprostaglandin dehydrogenase. Thus, an increase in PGE2 correlates with the increasing activity of P450c17.

Glucocorticoid treatment of sheep placental tissue specifically increases the rate of production of 17a,20a-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 17a-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 PGF2α production hours before the onset of uterine activity. A cause-and-effect relationship between the rise in estrogen and the appearance of PGF2α has been demonstrated in sheep. These events indicate that the decline in progesterone, the rise in estrogen, and the increase in PGF2α 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. In addition, there is a more extended time scale. Steroid changes in the sheep occur over the course of several days, while 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 steroids is avoided. In contrast to the maternal liver, the fetal liver has a limited capacity for transforming the biologically inactive cortisone.
Cortisol, in the presence of progesterone, stimulates (probably by blocking progesterone inhibition) trophoblastic CRH synthesis. Levels of the CRH-binding protein (also produced in trophoblast, membranes, and decidua) are decreased in the amniotic fluid and maternal circulation prior to

Increased CRH and decreased CRH-binding protein have been measured in women with preterm labor and in women with threatened preterm labor who

Arachidonic acid levels in the amniotic fluid also rise in labor, and arachidonate injected into the amniotic sac initiates parturition.

Prostaglandins induce labor.

Patients taking high doses of aspirin have a highly significant increase in the average length of gestation, incidence of postmaturity, and duration of labor.

Indomethacin prevents the normal onset of labor in monkeys and stops premature labor in human pregnancies.

During pregnancy, CRH levels in the amniotic fluid and the maternal circulation progressively increase, and although amniotic fluid levels do not further increase

The following observations support an important role for placental corticotropin-releasing hormone (CRH):

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 also 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. 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. Phospholipase A2 has been demonstrated in both human chorioamnion and uterine decidua. Although the precise mechanism for initiating prostaglandin synthesis, presumably by activation of the enzyme phospholipase A2, remains unknown, the availability of arachidonic acid for prostaglandin production during parturition follow the stimulation of hydrolysis of phosphatidylethanolamine and phosphatidylinositol in decidual, amnion, and chorion laeve tissues. Microsomes from amnion, chorion laeve, and decidua vera tissues contain lipases that hydrolyze fatty acids esterified in the 2 position. Specific phospholipase activity (phospholipase A2 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 formation 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 prostaglandine withdrawal in the activation of prostaglandin production cannot be discounted. 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 (DHAS), and this activity is increased around the time of parturition. In addition, the human fetal membranes synthesize and metabolize progesterone. The membranes contain a 17,20-hydroxysteroid dehydrogenase system. One active site converts 20a-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). 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. Throughout most of pregnancy, the amnion and chorion may exert an inhibitory influence over the myometrium by suppressing calcium channel activity. 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.
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 (also produced in trophoblast, membranes, and decidua) are decreased in the amniotic fluid and maternal circulation prior to labor. This decrease in the CRH-binding protein would allow an increase in CRH activity.
4. CRH stimulates prostaglandin release in fetal membranes, decidua, and myometrium.
5. 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.
6. Cortisol, in the presence of progesterone, stimulates (probably by blocking progesterone inhibition) trophoblastic CRH synthesis.
These observations are consistent with a key mechanism involving CRH activity in the initial triggering events of parturition. Although in the sheep the CRH signal begins in the fetal brain, in women it appears to begin in the uterus. Progesterone is the only major inhibiting factor for CRH production in placental tissues. It has been hypothesized that rising fetal cortisol secretion competes 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. The increase in CRH would augment fetal ACTH secretion, producing more fetal cortisol as well as more DHA to serve as precursor for the increase in estrogen that occurs prior to parturition; this indicates the important role of the fetal brain in the overall process.

The competitive inhibition of progesterone would in effect be a mechanism of progesterone withdrawal. This sequence of events could be started by an increase in CRH or a decrease in CRH-binding protein, or both. 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 relative hypoxemia. Regardless of the specific triggering event, it is increasingly clear that the fetus plays a pivotal, if not controlling, role in parturition.

### The Role 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 PGE₂ 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. The link between infection and the onset of labor (especially preterm labor) may be due to the conversion by bacterial medium (with factors such as the interleukins) of arachidonic metabolism in the membranes and decidua to a condition associated with labor marked by the production of PGE₂. In this case, prostaglandin production may be a consequence of inflammatory induction of the second cyclooxygenase enzyme, COX-2.

Undoubtedly, 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-a stimulate prostaglandin production by the fetal membranes apparently by regulating intracellular calcium concentrations. Secretory products of the fetal membranes themselves are active stimulators of membrane prostaglandin production, including renin derived from chorion prerenin. Decidual PGE₂ production is enhanced by bradykinin, epidermal growth factor and transforming growth-factor-a, and these responses are further increased by interleukin-1β. Prostaglandin production by amnion, chorion, and decidual cells is stimulated by corticotropin-releasing hormone and modulated by progesterone. The ubiquitous substances, activin and inhibin, are involved here as well. Amnion and chorion produce the activin and inhibit subunits, and activin stimulates prostaglandin PGE₂ release from amnion cells.

During labor the maternal circulating levels of PGE₂, PGF₁α and the PGF₂α-metabolite are increased, a change which can be directly attributed to uterine production in that 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. Meanwhile, prostacyclin and its metabolite are not increased.

### 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. Once labor has begun, oxytocin levels rise significantly, especially during the second stage. Thus, oxytocin may be 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. Part of the contribution of oxytocin to parturition is the stimulation of prostaglandin synthesis in decidua and myometrium. Cervical dilatation 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, chorion, and significantly, in the decidua. 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 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
Pharmacologically and physiologically, prostaglandins have two direct actions associated with labor: ripening of the cervix and myometrial stimulation. Successful labor 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 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.

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 and 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. 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 32–34 weeks. Perhaps it is the treatment of choice for the inhibition of labor during a maternal transplant. 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. Sulindac is just as effective as a tocolytics but does not affect urine output and amniotic fluid, and it has a lesser impact on the fetal ductus arteriosus.

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.

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<sub>2</sub> is very effective, while PGF<sub>2alpha</sub> 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<sub>2</sub> administered as tablets, suppositories, and mixed in gels has been very effective for cervical ripening. The commercial formulation currently available in the U.S. consists of 0.5 mg PGE<sub>2</sub> dissolved in a triacetin viscous gel base, also containing a colloid silicone dioxide, provided in a prefilled syringe for application. A synthetic PGE<sub>2</sub> analogue, misoprostol, is also safe and effective when used intravaginally for cervical ripening and labor induction.

A major clinical application for the induction of labor in the United States is the use of intravaginal PGE<sub>2</sub> in cases of fetal demise and anephric fetal. Based on our own experience and certain precautions have been developed. The patient should be well hydrated with an electrolyte solution to counteract the induced vasodilatation 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<sub>2</sub> is as effective as intravenous oxytocin, even including patients with previous cesarean sections. The intravaginal administration of misoprostol, the synthetic prostaglandin E<sub>1</sub> analogue, is safe, effective, and relatively inexpensive. 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, intra-amniotic prostaglandin, intramuscular methyl esters, and vaginal PGE suppositories are available. 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, antibiotic administration, readmission to the hospital, and more operative procedures when compared with saline abortions. The surgical procedure, D&E (dilation and evacuation) is safer and less expensive than medical methods, and it is better tolerated by patients.

The combination of prostaglandin's oxytocic action with the antiprogesterone effect of RU486 (mifepristone) has proved to be a safe and effective medical treatment for the induction of therapeutic abortion prior to 9 weeks gestation. An orally active methyl ester of prostaglandin E₁ is only 20% effective as an abortifacient when administered by itself. Combing the prostaglandin with RU486 achieves greater than 95% efficacy, safely and inexpensively.

**Prostaglandins and Postpartum Hemorrhage**

When routine methods of management for postpartum hemorrhage caused by uterine atony have failed, an analogue of prostaglandin F₂α has been used with excellent results (80–90% successful). Prostin 15 M (15-S)-15-methyl prostaglandin F₂α-15-bromethane. The dose is 0.25–0.5 mg, repeated up to 4 times and given with equal efficacy either intramurally 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. It should not be used in patients with severe hypertension or symptomatic asthma. Diarrhea is a frequent side effect.

**Prostaglandins and the Fetal Circulation**

The predominant effect of prostaglandins on the fetal and maternal cardiovascular system is to maintain the ductus arteriosus, 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 PGI₂ and TXA₂.

Fetal lamb ducts homogenates produce mainly PGI₂ when incubated with arachidonic acid. PGE₂ and PGF₂α are formed in small amounts and TXA₂ not at all. Although PGE₂ is less abundant than PGI₂ in the ductus, it is more potent vasodilator of the ductus and is more responsive to oxygen (decreasing vasodilatation with increasing oxygen). Thus, PGE₂ appears to be the most important prostaglandin in the ductus from a functional point of view, while PGI₂, 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 TXA₂ formation. This fits with the association of ducal 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, TXA₂ 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 patency after birth, while preparing an infant for surgery to correct a congenital lesion causing pulmonary hypertension. 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. 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. The highest incidence of successful indomethacin ductus closure has been with infants less than 30 weeks gestation and less 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. 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 which can be minimized if long-term use is avoided.

**Prostaglandins and Fetal Breathing**

Prior to parturition, fetal breathing is very shallow. It is proposed that placental PGE₂ suppresses breathing by acting in the fetal brain. Occlusion of the umbilical cord is rapidly followed by a loss of this PGE₂ influence and the onset of air breathing. The administration of indomethacin to fetal sheep increases, while infusion of PGE₂ 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, palmitylmyristoyl 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 dipalmityl 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 were the first to demonstrate that the L/S ratio correlates with pulmonary maturity of the fetal lung. 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. 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 may actually occur within 3–4 days.

An increase in the surfactant content of phosphatidylglycerol 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 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.

Abnormalities of pregnancies 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 associated with higher fetal weight gain, higher birthweight, and higher gestational age is associated with better outcomes. Poorer outcomes occur when PG is not detectable in the amniotic fluid. Accelerated lung 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 Liggin observed survival of premature lambs following the administration of cortisol to the fetus, 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 that reason, fetal lung maturation can best be viewed as the result of not only cortisol, but the synergistic action of cortisol, prolactin, thyroxine, estrogens, and factors in amniotic fluid. Growth hormone and thyrotropin-releasing hormone (TRH) also play a role in fetal lung maturation. TRH is 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.

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. The clinician should always have a high index of suspicion for thyroid dysfunction because of the influence of the hypothalamic-pituitary-thyroid axis in the postpartum period. This is a high-risk time for the development of thyroid dysfunction.


doi:10.1210/jc.117.8.884


Carr BR, Simpson ER, Umilical cord plasma levels of dehydroepiandrosterone sulfate during human gestation, Endocrinology 1983;112:884–890

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 supression 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, like thyroiditis.

Chapter References


Abnormalities of sexual differentiation are seen infrequently in an individual clinician's practice. There are, however, few practitioners who have not been challenged at least once by a newborn with ambiguous genitalia or by a young woman with primary amenorrhea on a genetic basis. The categorization of the various syndromes in this area has been confusing, requiring constant reference to multiple textbooks, and dependence upon memory of eponym-laden, seemingly endless lists of syndromes. Happily, this "catalogue" state of affairs has changed; major advances in reproductive science have yielded clarification and consolidation. As a result, an informed basis for clinical practice has emerged and is readily applicable.

This chapter will present classification of the major problems and our clinical approach to diagnosis. Normal sexual differentiation will be considered in order to provide a basis of understanding for the various types of abnormal development. This is followed by a section on the diagnosis and management of ambiguous genitalia. Some subjects are discussed in other chapters, but brief descriptions will be repeated here in order to present a complete picture. It will be seen that analysis of phenotypic ambiguity follows a fundamental, pervasive theme: too little androgen effect in males, too much androgen effect in females. Whereas androgen biologic "availability" may be excessive in females because of abnormally high levels of intake or production, reduced androgen effects in males can be the result of defects in synthesis, peripheral and target organ conversion, abnormalities in androgen receptor or receptor-DNA interactions, as well as defective gonadal development and function.

Normal Sexual Differentiation

The gender identity of a person (whether an individual identifies as a male or a female) is the end result of genetic, hormonal, and morphologic sex as influenced by the environment of the individual. It includes all behavior with any sexual connotation, such as body gestures and mannerisms, habits of speech, recreational preferences, and content of dreams. Sexual expression, both homosexual and heterosexual, can be regarded as the result of all influences on the individual, both prenatal and postnatal. Specifically, gender identity is the result of the following determinants: genetic sex, gonadal sex, the internal genitalia, the external genitalia, the secondary sexual characteristics that appear at puberty, and the role assigned by society in response to all of these developmental manifestations of sex.

Prenatally, sexual differentiation follows a specific sequence of events. First is the establishment of the genetic sex. Second, under the control of the genetic sex the gonads differentiate, determining the hormonal environment of the embryo, the differentiation of internal duct systems, and the formation of the external genitalia. It has become apparent that the embryonic brain is also sexually differentiated, perhaps via a control mechanism very similar to that which determines the sexual development of the external genitalia. The inductive influences of hormones on the central nervous system may have an effect on the patterns of hormone secretion.
and sexual behavior in the adult.\textsuperscript{1, 3–5, 9, and 6}

**Gonadal Differentiation**

In human embryos, the gonads begin development during the 5th week of gestation as protuberances overlying the mesonephric ducts. The migration of primordial germ cells into these gonadal ridges occurs between weeks 4 and 6 of gestation. Although germ cells do not induce gonadal development, if the germ cells fail to arrive, gonads do not develop and only the fibrous streak of gonadal agenesis will exist (Chapter 3). At 6 weeks of gestation (4 weeks after ovulation) the gonads are indifferent but bipotential, possessing both cortical and medullary areas, and are capable of differentiation into either testes or ovaries. They are composed of germ cells, epithelial mesenchyme (potential granulosa/Sertoli cells), mesenchyme (potential theca/Leydig cells), and the mesonephric duct system. Wolffian and müllerian ducts exist side by side; external genitalia are undifferentiated. Subsequent sexual differentiation requires direction by various genes, with a single gene determinant on the Y chromosome (testes determining factor — TDF) necessary for testicular differentiation, beginning at 6–7 weeks gestation.\textsuperscript{4}

Since the identification of the Y chromosome's importance to male differentiation over 4 decades ago, three proteins have been suggested as the Y-encoded, gene expressed, testes determining factor. The first was the H-Y histocompatibility antigen and the second, ZFY (a zinc finger protein). Both were abandoned because of inconsistencies of expression in various cell types (in XX males and XY females), as well as absent expression in indistinguishable males with testes. More recently, SRY (Sex determining Region Y) has been isolated, almost certainly the true sex-determining region on the short arm of the Y chromosome, the only gene on the Y chromosome required for sex determination.\textsuperscript{3}

SRY is a single copy gene located in the smallest Y chromosome region capable of sex reversal. It is expressed in the genital ridge only during the appropriate time of embryonic development when testicular cords form; it is deleted or mutated in cases of human XY females, and present in 46,XX males, and it can sex reverse XX mice into males.\textsuperscript{11 and 12} The 204 amino acid protein product contains an 80 amino acid domain with a motif shared by a recognized family of transcription factors (the high mobility group) that bind to DNA and regulate gene transcription. Investigations of the DNA binding properties of the SRY high mobility group (HMG) box in the promoter regions of P450 aromatase (conversion of testosterone to estradiol which is down-regulated in the male embryo) and anti-müllerian hormone (responsible for regression of the müllerian ducts) support the hypothesis that SRY directly controls male development through sequence-specific regulation of target genes.\textsuperscript{12}

SRY participation in morphogenesis leading to a testis from the bipotential genital ridge is a model "genetic switch" between alternative inherent programs. Whereas testes formation is an active event, female sex determination is the default pathway occurring if SRY is absent or deficient. Genes other than SRY are also required for proper gonadogenesis.\textsuperscript{13, 14} In the human, autosomal genes are essential for gonadal development. These autosomal genes regulate migration of the germ cells and coding of the steroidogenic enzymes. The formation of the testicle precedes any other sexual development in time, and a functionally active testis controls subsequent sexual development, therefore SRY presumably controls these autosomal genes. Thus, testicular hormones activate or repress genes to direct development away from an otherwise predetermined course of female differentiation.

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.\textsuperscript{14, 15} SF-1 is also believed to be involved in the production of anti-müllerian hormone.\textsuperscript{16} 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.\textsuperscript{17}

The WT1 gene is named after the Wilms' tumor nephroblastoma because it is one of the genes on chromosome 11 deleted in patients with this tumor. Mutant mice lacking WT1 fail to develop kidneys and gonads. WT1 mutations, however, could not be detected in 25 patients with a congenital absence of the uterus and vagina, indicating that WT1 may be necessary for normal renal and gonadal development, but not for early müllerian duct development.\textsuperscript{18}

Testicular differentiation begins at 6–7 weeks; first with Sertoli cells that aggregate to form spermatogenic cords, then seminiferous tubules, followed by Leydig cell formation a week later. Human choric gonadotropin (HCG) stimulation produces Leydig cell hypertrophy, and peak fetal testosterone levels are seen at 15–18 weeks.\textsuperscript{19} It is very probable, but not conclusively proven, that HCG stimulates steroidogenesis in the early fetal testes, so that androgen production will ensue and masculinization can be accomplished.\textsuperscript{20}

In an XX individual, without the active influence of a Y chromosome, the bipotential gonad develops into an ovary about 2 weeks later than testicular development. The cortical zone develops and contains the germ cells, while the medullary portion regresses with its remnant being the rete ovarii, a compressed nest of tubules and Leydig cells in the hilus of the ovary. The germ cells proliferate by mitosis, reaching a peak of 5–7 million by 20 weeks. By 20 weeks, the fetal ovary achieves mature compartmentalization with primordial follicles containing oocytes, initial evidence of follicle maturation and atresia, and an incipient stroma. Degeneration (atresia) begins even earlier, and by birth, approximately 1–2 million germ cells remain. These have become surrounded by a layer of follicular cells, forming primordial follicles with oocytes that have entered the first meiotic division until reactivation of follicular growth that may not occur until years later. Excessively rapid atresia (germ cell attrition) in gonadal dysgenesis (45,X) accounts for the streak gonad seen in these cases.\textsuperscript{21} A complete 46XX chromosomal complement is necessary for normal ovarian development.\textsuperscript{22} The second X chromosome, therefore, contains elements essential for ovarian development and maintenance.

**Duct System Differentiation**

Caspar Wolff described the mesonephros in 1759 in his doctoral dissertation when he was 26 years old.\textsuperscript{23} The paired structures of the mesonephros of the early vertebrate embryo were named wolffian bodies by the 19th century embryologist, Rathke, in recognition of Wolff's initial discovery and description. Johannes Müller, a German physiologist with a prodigious academic output, described the embryology of the genitalia in 1830. The paramesonephric ducts received his name, not because of his original contributions, but because of his ability to synthesize current knowledge in his effective writings. His physiology text was a standard in many

\[ \text{Diagram} \]
Renal development goes through 3 stages: pronephric, mesonephric, and metanephric. The mesonephric ducts remain as internal genitalia. At this stage the mesonephric ducts are called the wolffian ducts. The paired paramesonephric ducts are the müllerian ducts. The wolffian and müllerian ducts are discrete primordia which 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 3rd fetal month, except for nonfunctional vestiges.

Hormonal control of mammalian somatic sex differentiation was established by the classic experiments of Alfred Jost. In Jost’s landmark studies, the active role of male determining factors 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 perhaps even the brain. The critical factors in determining which of the duct structures stabilize or regress are the secretions from the testes: testosterone and anti-müllerian hormone (AMH), also known as müllerian inhibiting substance or müllerian inhibiting factor.

AMH is a member of the transforming growth factor-β family of glycoprotein differentiation factors that includes inhibin and activin. The gene for AMH has been mapped to the short arm of chromosome 19, and the AMH receptor gene is on chromosome 12. AMH is synthesized by Sertoli cells (activated by SRY) soon after testicular differentiation and is responsible for the ipsilateral regression of the müllerian ducts by 8 weeks, before the emergence of testosterone and stimulation of the wolffian ducts. 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. Knockout mice for AMH develop normal testes and ovaries. 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, abnormalities in the renal system are associated with abnormalities in the development of the tubes, uterus, and upper vagina.

AMH has extra müllerian functions. AMH exerts an inhibitory effect on oocyte meiosis, plays a role in the descent of the testes, and inhibits surfactant accumulation in the lungs. Proteolytic cleavage of AMH produces fragments which have the ability to inhibit growth of various tumors (a potential therapeutic application). Testicular descent occurs in stages. Transabdominal movement of the testes is the result of rapid gubernacular growth, apparently under AMH control. Movement through the inguinal canal is mediated by androgens.

AMH, secreted by the Sertoli cells, is detectable in the serum of males during infancy, childhood, adolescence, and adulthood (with a decline to barely detectable levels after puberty). In contrast, AMH, secreted by granulosa cells, is not measurable until puberty in females. This difference allows serum measurement to be a sensitive marker for the presence of testicular tissue in intersex anomalies. After puberty, testosterone, with an added contribution from meiotic germ cells, suppresses AMH secretion in males, and, therefore, individuals with the androgen insensitivity syndrome (androgen receptor defects) have very high AMH levels after puberty. The failure of testosterone to suppress AMH secretion during fetal and newborn life is explained by the absence of the androgen receptor in Sertoli cells until later in life.

Testosterone is secreted by the fetal testes soon after Leydig cell formation (at 8 weeks) and rapidly rises to peak concentrations at 15–18 weeks. This testosterone secretion stimulates development of the wolffian duct system into epididymis, vas deferens, and seminal vesicles. Testosterone levels in the male fetus correlate with Leydig cell development, overall gonadal weight, 3b-hydroxysteroid dehydrogenase activity, and chorionic gonadotropin (HCG) concentrations. As HCG declines (approximately 20 weeks) the fetal pituitary luteinizing hormone (LH) assumes control of Leydig cell testosterone secretion; amenorrhoeas and other forms of congenital hypopituitarism display diminished androgen effects on internal and external genitalia. Dihydrotestosterone (DHT) is the active form of testosterone. It is metabolized inside the target cells of the wolffian ducts and gives rise to special ducts and glands, whereas the other disappears during the 3rd fetal month, except for nonfunctional vestiges.

The wolffian ducts receive testosterone signals directly from nearby Leydig cells as well as the general fetal circulation. This local paracrine effect is essential to the stimulation of ipsilateral differentiation into the epididymis, vas deferens, and seminal vesicles. Duct system differentiation will proceed, therefore, according to the nature of the adjacent gonad. The wolffian ducts do not form dihydrotestosterone, so the direct high concentration is crucial for normal development. Because of this local paracrine action, wolffian development cannot be stimulated in females exposed to adrenal or exogenous androgens.

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. These classic roles for the presence and absence of testosterone are undisputed; however, estrogen may make a contribution in this process. Knockout mice for the estrogen receptor-a have a marked retention of wolffian tubules, indicating that estrogen plays a role in the degeneration of the wolffian system.

External Genitalia Differentiation

In the bipotential state (6th gestational week), the external genitalia consist of a genital tubercle, a urogenital sinus, and two lateral labioscrotal 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. Normally, this differentiation is under the active influence of androgens from the Leydig cells of the testis. The genital tubercle forms the penis, labioscrotal folds fuse to form a scrotum, and folds of the urogenital sinus form the penis urethra. The testis begins androgen secretion by 8–9 weeks; masculinization of the external genitalia is manifest 1 week later and is completed by 14 weeks. To achieve this morphologic change, external genitalia target tissue cells must convert testosterone to dihydrotestosterone (DHT) by the intracellular enzyme 5a-reductase. In the male, DHT mediates the following androgen events: temporal hairline recession, growth of facial and body hair, development of acne, and development of the external genitalia and prostate.
In the absence of this androgen effect (the absence of a Y chromosome, the presence of an ovary, the absence of a gonad, abnormalities in androgen receptor or postreceptor events, or defects of the 5α-reductase enzyme), the folds of the urogenital sinus remain open, forming the labia minora, the labioscrotal folds form the labia majora, the genital tubercle forms the clitoris, and the urogenital sinus differentiates into the vagina and the urethra. Thus, the lower vagina is formed as part of the external genitalia.

Exposure to androgens at critical time periods leads to variable masculinization. Androgen exposure at 9–14 weeks superimposes variable external ambiguity on the basic female phenotype (clitoral hypertrophy, hypospadias, scrotalization of nonfused labia). By the same token, incompletely masculinized genitalia will result if sufficient local androgen concentration or activity is not achieved by the 12th week in the male. Because of shared common tissue origin, male-female external genital structural ambiguities reflect abnormal androgen impact: males too little, females too much.

Central Nervous System Differentiation

At the same time the presence or absence of androgens is playing a critical role in genitalia development, the neuroendocrine mechanism of the central nervous system is also being influenced. Androgens present in sufficient amounts during the appropriate critical stage of development may program the central nervous system (CNS) to induce the potential for male sexual behavior. Experimental and analytical evidence suggests that a behavioral effect can be traced to this early androgen influence. Inappropriate fetal hormonal programming may contribute, therefore, to the spectrum of psychosexual behavior seen in humans. In addition, gender role is heavily influenced by assignment of sex of rearing followed by social interaction based upon genital appearance and the development of secondary sexual characteristics.

Abnormal Sexual Differentiation

The standard classification of individuals with intersexuality (hermaphroditism) proceeds according to gonadal morphology. In this terminology, a true hermaphrodite possesses both ovarian and testicular tissue. A male pseudohermaphrodite has testes, but external and sometimes internal genitalia take on female phenotypic aspects. A female pseudohermaphrodite has ovaries, but genital development displays masculine characteristics. These classifications are modified to reflect gonadal abnormalities due to abnormal sex chromosome constitution or abnormalities of phenotype attributable to an inappropriate fetal hormone environment. Hypospadias in the absence of any other deformity is not included in this classification.

Disorders of Fetal Endocrinology

Masculinized females (female pseudohermaphroditism)

Congenital adrenal hyperplasia

21-Hydroxylase (P450c21) deficiency

11b-Hydroxylase (P450c11b) deficiency

3b-Hydroxysteroid dehydrogenase deficiency

Elevated androgens in the maternal circulation

Drug intake

Maternal disease

Aromatase (P450arom) deficiency

Incompletely masculinized males (male pseudohermaphroditism)

Androgen insensitivity syndromes

5α-Reductase deficiency

Testosterone biosynthesis defects
A variety of mutations affecting CYP21B lead to 21-hydroxylase deficiency, but 85% are gene conversions. Most patients are compound heterozygotes, having a
Two 21-hydroxylase genes exist, designated CYP21A and CYP21B, located on chromosome 6, in tandem duplication with the genes encoding the fourth
Developments in molecular biology and genetics have greatly expanded our understanding of this condition. The syndrome may appear in utero or develop postnataally.
Because the female external genitalia phenotype is not completed until 20 weeks of fetal age, early androgen excess (10–12 weeks gestation) may fully masculinize, whereas late (18–20 weeks gestation) androgen can create limited ambiguity of the basically female appearance of the urogenital sinus and genital folds. The size of the clitoris depends on the quantity rather than timing of androgen excess. Cases of incorrect sex assignment in the female are due to the similarity between these external genitalia and hypospasias and bilateral cryptorchidism in a male infant.
If untreated, the female with adrenal hyperplasia will develop signs of progressive virilization postnataally. Pubic hair will appear by age 2–4, followed by axillary hair, then body hair and beard. Bone age is advanced by age 2, and because of early epiphyseal closure, height in childhood is achieved at the expense of shortened stature in adulthood. Progressive masculinization continues with the development of the male habitus, acne, deepened voice, and primary amenorrhea and infertility.
In addition to sexual changes, patients can present with metabolic disorders such as salt-wasting, hypertension, or rarely, hypoglycemia. An electrolyte imbalance of the salt-losing type is usually apparent within a few days of birth and occurs in approximately two-thirds of patients with virilizing adrenal hyperplasia. A salt-losing crisis usually begins 5–15 days after birth. Beginning with a refusal to feed, failure to thrive, apathy, and vomiting, the infant goes on to an Addisonian-like crisis with hyponatremia, hyperkalemia, and acidosis. Rapid diagnosis and treatment are necessary to save these infants. Less frequent is hypertension, which occurs in approximately 5% of patients with virilizing adrenal hyperplasia.
Viritizing adrenal hyperplasia is the result of an inherited abnormality of steroid biosynthesis that 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. Therefore, one can see a specific enzyme deficiency is due to effects of both the inadequate production of cortisol/aldosterone and excess accumulation of precursors, with diversion into biosynthetic pathways yielding androgens.

The most common enzymatic defects are the 21-hydroxylase (P450c21), the 11b-hydroxylase (P450c11), and the 3b-hydroxysteroid dehydrogenase types. Very rarely, blocked synthesis of cortisol can be due to a defect in P450c17.

1. The disorder is inherited as a monogenic autosomal recessive trait.
2. Two 21-hydroxylase genes exist, designated CYP21A and CYP21B, located on chromosome 6, in tandem duplication with the genes encoding the fourth component of complement. Only CYP21B is active in adrenal steroidogenesis; CYP21A is not involved (a pseudogene, also known as CYP21P, because it is inactive).
3. A variety of mutations affecting CYP21B lead to 21-hydroxylase deficiency, but 85% are gene conversions. Most patients are compound heterozygotes, having a different genetic lesion on each copy of chromosome 6, one from each parent. The severity of the condition is determined by the activity of the least affected
congenital adrenal hyperplasia are due to 11b-hydroxylase deficiency, about 1 in 100,000 births, with a higher incidence in Jewish people who originated in Morocco. 1

Usually as a result of 11b-hydroxylase deficiency, metabolically active precursors of cortisol and cortisone add to excess androgen synthesis as further liabilities of ACTH-induced hyperplasia. Hypertension and hypokalemic alkalosis are induced by elevated deoxycorticosterone with reduced renin and aldosterone. Virilization is caused by androgens of the “deoxy” type (dehydroepiandrosterone, dehydroepiandrosterone sulfate, and androstenedione). The diagnosis is confirmed by high plasma deoxycorticosterone and compound S (11-deoxycortisol) levels.

About two-thirds of untreated patients with 11b-hydroxylase deficiency become hypertensive, usually of mild to moderate degree (150/90 mm Hg) and only after several years of life. A mild nonclassic form of 11b-hydroxylase deficiency, as in 21-hydroxylase defects, has also been documented; it is characterized by mild biochemical abnormalities, and the patients are only slightly virilized and rarely hypertensive.

Contrary to 21-hydroxylase deficiency, the 11b-hydroxylase deficiency locus is remote from the HLA complex. The gene for the enzyme is on the long arm of chromosome 8, and the deficiency is inherited in autosomal recessive fashion. Actually there are two isozymes with 11b-hydroxylase activity, encoded by two genes on chromosome 8. One, CYP11B1 or P450c11, is the enzyme classically referred to as 11b-hydroxylase and regulated by ACTH. The other, CYP11B2 or P450c18 (or P450aldo), has been called aldosterone synthase and is regulated by angiotensin II. P450c18 has strong 11b-hydroxylase activity, but also hydroxylates and oxidizes at the C-18 position. Deficiency in 11b-hydroxylase results from mutations in P450c11 (CYP11B1).

Enzyme Defects in Adrenal and Ovary: Deficient 3β-Hydroxysteroid Dehydrogenase. Lack of this essential step in the formation of all biologically active steroids affects both the adrenal cortex and the ovary and is also inherited in autosomal recessive fashion. Thus, there is decreased synthesis of glucocorticoids, mineralocorticoids, androgens, and estrogens. These infants are severely ill at birth and rarely survive. The external genitalia ambiguity results from the massive increase in dehydroepiandrosterone that is androgenic when available in excess, and also can be utilized to form more potent androgens in peripheral tissues. Thus, females may be slightly virilized and males incompletely masculinized with a variable degree of hypospadias. The spectrum of clinical phenotypes also includes both salt wasting and non-salt wasting forms. The degree of the enzyme defect cannot be extrapolated from the degree of external genitalia ambiguity. As in 21-hydroxylase deficiency, milder nonclassic cases may occur with mild hirsutism and elevated dehydroepiandrosterone and dehydroepiandrosterone sulfate being the only distinguishing features. Some argue that late-onset 3β-hydroxysteroid dehydrogenase deficiency is encountered more frequently than 21-hydroxylase deficiency. However, although an exaggerated 17a-hydroxyprogrenolone response to ACTH stimulation is common in women with hyperandrogenism, the response is consistent with adrenal hyperactivity and not an enzyme deficiency.

Furthermore, molecular studies fail to find mutations in the genes for the two 3β-hydroxysteroid dehydrogenase enzymes in patients with apparent mild to moderate 3β-hydroxysteroid dehydrogenase deficiency.

Enzyme Defect in Adrenal and Ovary: Deficient 17α-Hydroxylase (P450c17). With block of the 17α-hydroxylase enzyme (P450c17), synthesis of cortisol, androgens, and estrogens is curtailed. Only the non-17-hydroxylated corticoids, deoxycorticosterone and corticosterone, are formed. The molecular basis for this enzyme deficiency is due to a variety of mutations which result in multiple base deletions and duplications in the CYP17 gene on chromosome 10q24–25. The resulting syndrome is composed of hypertension (due to hypernatremia and hypervolemia), hypokalemia, infantile female external genitalia, which do not mature at puberty, and primary amenorrhea with elevated follicle-stimulating hormone (FSH) and luteinizing hormone (LH). Genital ambiguity is a problem only in male infants.

Epidemiology

Only the 21-hydroxylase deficiency has been studied extensively, in part because it is not only the most frequent cause of genital ambiguity and congenital adrenal hyperplasia but also because of the high prevalence of the nonclassical, late onset forms of the disease. Within families, the clinical picture is uniform, the type of syndrome (simple, salt-wasting, hypertensive) is usually, but not always, the same in affected siblings. The genetic defect in virilizing adrenal hyperplasia is an autosomal recessive gene. The ratio in offspring of unaffected parents is one affected to three nonaffected individuals. Treated patients have a 1:100 to 1:200 chance of producing an affected infant. Males and females are at equal risk. The classic form is a relatively common inborn error of metabolism. One out of every 100 Caucasians is likely to be a genetic carrier of the classic type, and neonatal screening tests indicate an incidence of 1 in 14,000 births. The highest frequency for congenital adrenal hyperplasia is in Alaskan Yupik Eskimos.

For the nonclassical types, frequency rates established by the usual methodology (neonatal screening, case surveys) are likely to markedly underestimate what may be one of the most common autosomal recessive disorders in humans. Extrapolations from ACTH testing suggest the following frequency.

<table>
<thead>
<tr>
<th>Nonclassical Disease</th>
<th>Heterozygous Carrier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Europeans</td>
<td>1 in 30</td>
</tr>
<tr>
<td>Hispanics</td>
<td>1 in 40</td>
</tr>
<tr>
<td>Slavs</td>
<td>1 in 50</td>
</tr>
</tbody>
</table>
Prenatal Diagnosis

The diagnosis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency can be obtained prenatally by demonstrating elevated levels of 17-hydroxyprogesterone (17-OHP), 21-deoxycortisol, and androstenedione in the amniotic fluid. 17-OHP may be elevated only in the salt-losing form of adrenal hyperplasia, but androstenedione is usually increased with all forms. The 11b-hydroxylase deficiency is associated with elevated levels of 11-deoxycortisol in amniotic fluid and tetrahydro-11-deoxycortisol in maternal urine. Because amniotic fluid steroids may be within the normal range with less severe deficiencies, hormonal measurements have been replaced by molecular genetic diagnosis.

Prenatal diagnosis of the 21-hydroxylase deficiency by chorion villus biopsy utilizing DNA probes offers the timely options of termination or in utero therapy. With chorion villus biopsy, diagnosis can be made and therapy instituted before the critical period of fetal genital differentiation with avoidance of genital ambiguity in affected female fetuses. In addition, masculinization of the fetal brain can be avoided which might have an impact on gender identity and adult sexual behavior. 

Despite the fact that prenatal diagnosis is not 100% accurate, prenatal treatment beginning at 4–5 weeks gestation has been administered with dexamethasone in fetuses at risk for 21-hydroxylase deficiency. Using multiple daily doses of dexamethasone (total no greater than 1.5 mg/day), complete prevention has been achieved in some newborns and diminished virilization in others. No congenital malformations or low birth weight or height have resulted from pregnancy-long cortisolloid derivative therapy. However, this treatment is associated with significant maternal side effects, such as severe striae with permanent scarring, hyperglycemia, hypertension, gastrointestinal symptoms, and emotional liability. A reduction in dosage during the second half of pregnancy is recommended; dosage can be titrated by maintaining the maternal serum estril levels in the normal range.

Diagnosis of Adrenal Hyperplasia

For years the demonstration of a metabolic defect and its location depended upon the study of urinary steroid excretion. Today, the immunassay of blood 17-hydroxyprogesterone (17-OHP) has become the primary assessment for the diagnosis and management of congenital adrenal hyperplasia. With the 21-hydroxylase and 11b-hydroxylase deficiencies, the 17-OHP level will be 50–400-fold above normal. In addition, plasma renin activity is measured to assess the degree of mineralocorticoid deficiency.

During delivery of normal infants, the concentration of 17-OHP is elevated in cord blood (1000–3000 ng/dL), but it rapidly decreases to 100–200 ng/dL after 24 hours. A delay in measurement gains accuracy. In contrast to 17-ketosteroids in the urine where the delay must be several days, with 17-OHP the delay need be only a day or two. In affected infants, 17-OHP ranges from 3000 to 40,000 ng/dL. Measurement of 17-OHP is the basis for the newborn screening programs currently in place in many countries and some states in the U.S. In Texas, the screening program detected classic congenital adrenal hyperplasia in 1 in 16,000 births, with a ratio of salt-wasting to simple virilizing of 2.7:1.

In adults, 17-OHP must be measured first thing in the morning to avoid later elevations due to the diurnal pattern of ACTH secretion. The baseline 17-OHP level should be less than 200 ng/dL. Levels greater than 200 ng/dL, but less than 800 ng/dL, require ACTH testing (discussed in Chapter 13). Levels over 800 ng/dL are virtually diagnostic of the 21-hydroxylase deficiency. The dehydroepiandrosterone sulfate level is usually normal. The hallmarks of nonclassical adrenal hyperplasia are elevated levels of 17-OHP and a dramatic increase after ACTH stimulation. The elevated levels of 17-OHP are often not impressve (e.g., overlapping with those found in women with polycystic ovaries due to anovulation), and a simple ACTH stimulation test must be utilized. To distinguish the 21-hydroxylase deficiency from other disorders, all of the following in addition to 17-OHP should be measured at zero and 60 minutes after ACTH stimulation: pregnenolone, 17-hydroxyprogrenolone, dehydroepiandrosterone, 11-deoxycortisol, cortisol, and testosterone.

Of course, in patients with 3b-hydroxysteroid dehydrogenase or 17-hydroxylase block, the blood levels of DHA and DHA sulfate (DHAS) will be markedly increased. In the 11b-hydroxylase deficiency, in addition to elevated 17-OHP, elevation of 11-deoxycortisol is diagnostic. In this deficiency, plasma renin activity will be low, whereas in 21-hydroxylase and 3b-hydroxysteroid dehydrogenase deficiencies plasma renin activity is elevated in the salt losing forms.

Treatment

Treatment of adrenal hyperplasia is to supply the deficient hormone, cortisol. This decreases ACTH secretion and lowers production of androgen precursors. The addition of salt-retaining hormone to glucocorticoid therapy has improved the control of the disease. When the plasma renin activity is normalized, ACTH and androgen levels are further decreased, and a decrease in the glucocorticoid dose is also possible. Therefore, the modern management of hormonal control requires the measurement of the blood levels of 17-OHP, androstenedione, testosterone, and plasma renin activity.

The drugs of choice are hydrocortisone (cortisol), approximately 10 mg per day, and 9-fluorohydrocortisone, approximately 100 mg/day. This method of treatment monitoring applies to all forms of adrenal hyperplasia. The standard dose of cortisol is 12–18 mg/m² or 3.5–5 mg/m² of prednisone, but larger doses given on alternate days (14 mg/m² of prednisone, about 20 mg) can maintain adrenal androgen suppression and perhaps achieve better growth and pubertal development, despite higher levels of 17-hydroxyprogesterone. The 17-OHP level should be maintained in the range of 500 to 4000 ng/dL, thereby avoiding both overtreatment and undertreatment. Minor stresses will cause brief elevations of adrenal androgens but usually do not require readjustment of dosage. With major stress, such as surgery, additional hormonal support is necessary.

The surgical treatment of the anatomical abnormalities should be carried out in the first few years of life, when the patient is still too young to remember the procedure and too young to have developed psychological problems centered about the abnormal external genitalia. If clitoridectomy is necessary, the clitoral recession procedure, conserving the glans and its innervation, should be employed. It is important to know that women who undergo total clitoral amputations have no normal sexual function. Therefore, clitoridectomy should be avoided unless the clitoris is too large to permit passage of the baby.
promptly corrected with glucocorticoid therapy.⁶⁵ Consistent with the belief that androgens can program sexual behavior during fetal development, women with congenital adrenal hyperplasia report a decrease in heterosexual activity and an increase in homosexual activity.⁶⁶

The maintenance steroid dose usually does not need to be changed during pregnancy. The dosage of steroids used in the treatment of this syndrome replaces the approximate amount normally produced and, therefore, is a physiologic dose. At these low doses, teratogenic effects would be unlikely, and none have been noted. The need for additional steroids during the stress of labor and delivery is obvious and is usually met by the administration of cortisone acetate intramuscularly and cortisol intravenously. Infection and impaired wound healing have not been problems. Aside from the liability associated with genetic transmission of this syndrome, the children born to patients with adrenal hyperplasia have been normal. The newborn should be closely observed for adrenal insufficiency due to steroid crossover and suppression of the fetal adrenal in utero.

**Treatment Problems**

Overtreatment causes Cushing’s syndrome and poor growth; undertreatment is associated with short stature, hirsutism, and infertility. In some cases, undertreatment and increased androgen secretion lead to premature pubertal maturation that may require treatment with a gonadotropin-releasing hormone (GnRH) agonist. The adult height achieved by most patients is less than normal, testimony to overtreatment and undertreatment (which both compromise growth). Mineralocorticoid therapy should be maximized (maintaining the plasma renin activity at its lower limit of normal) to eliminate hypovolemia as a stimulus for ACTH secretion, via the angiotensin II pathway.

Because it is now recognized that estrogen is the primary regulator of bone response in both males and females, low-dose hydrocortisone and fludrocortisone treatment was combined with flutamide (an antiandrogen) and testolactone (an inhibitor of androgen to estrogen aromatization). In a preliminary study, normalization of growth was achieved.⁶⁶

**Masculinization Due to Elevated Androgens in the Maternal Circulation**

Masculinization of the female fetus, although in most cases due to fetal virilizing adrenal hyperplasia, can be produced by an androgen-secreting maternal tumor or can be due to the intake of exogenous androgenic substances, such as progestins and danazol. When not caused by an error in the metabolism of the fetal adrenal gland, virilization is not progressive, blood steroids are not elevated, and no hormonal therapy is needed. Subsequent development will be normal. Therefore, surgical correction of abnormalities in the external genitalia is the only indicated treatment.

The occurrence of an androgen-secreting tumor in a mother during pregnancy is rarely seen. On the other hand, the iatrogenic cause of masculinization is a well known story. The majority of these cases resulted from antenatal maternal treatment of threatened or recurrent miscarriage with various progestin compounds. In view of the lack of evidence for positive results with such therapy, progestin compounds, other than progesterone, are no longer administered to pregnant women.

**Aromatase (P450arom) Deficiency**

P450 arom is encoded by the CYP19 gene on chromosome 15p21.1. A deficiency in this enzyme is very uncommon, and only rare mutations in P450arom have been reported.⁶⁷ An individual with this deficiency will be first affected by the failure to achieve normal levels of aromatization within the placenta. A placental deficiency in aromatase activity would allow an accumulation of the fetal androgen precursors utilized in placental estrogen synthesis. This condition is associated with virilization of the mother during the second half of pregnancy, low estrogen levels in the mother, and a female newborn with masculinization. Accurate prenatal diagnosis requires a loading test with DHA and DHAS. A patient with a placental sulfatase deficiency will increase her estrogen levels in response to DHA and not to DHAS. A patient with an aromatase deficiency will respond to neither steroid. Pubertal development is prevented because the ovary cannot aromatize androgens in order to produce estrogen; a patient can present with primary amenorrhea (hypergonadotropic hypogonadism) and mild virilization.

**Incompletely Masculinized Males**

Incompletely masculinized males are male by genetic sex (XY) and possess testicles, but the external genitalia are not normally male. Male pseudohermaphrodites may arise in one of four ways:

1. Defective responses in androgen dependent tissues — Androgen Insensitivity Syndromes.
2. Abnormal androgen synthesis.
3. Gonadotropin resistant testes.
4. Absent or defective anti-müllerian hormone.

** Syndromes of Androgen Insensitivity**

Factors that influence the response to androgens in specific target cells include the following:

1. The intracellular concentration of androgen.
2. The relative binding affinity of these steroids to their nuclear androgen receptors.
3. The binding capacity of the receptor.
4. The nuclear content of androgen receptors.
5. The cellular concentrations of catabolic and/or synthetic enzymes (e.g., 5a-reductase, aromatase, 17b-hydroxysteroid dehydrogenase).
6. The adequacy of the nuclear (chromatin) acceptor site.
7. The adequacy of regulatory molecules (adapter proteins) controlling chromatin “read” of the androgen message.
8. RNA processing and translation.

Defects in androgenization of targets theoretically can occur as a result of failure in any of these steps. Three major clinical conditions are worthy of detailed review.
Complete Androgen Insensitivity — Testicular Feminization

Complete androgen insensitivity was first described in detail by Morris at Yale, who provided the descriptive term, testicular feminization. The phenotype of this condition (also discussed in Chapter 11) is female, despite the normal male karyotype, 46,XY. There is a congenital insensitivity to androgens, transmitted by means of a maternal X-linked recessive gene responsible for the androgen intracellular receptor. Therefore, androgen induction of Wolffian duct development does not occur. However, anti-müllerian hormone activity is present, and the individual does not have müllerian development (a natural experiment that indicates the presence of an anti-müllerian hormone). Frequently the testes have descended to the inguinal canal because AMH mediates the transabdominal descent of the testes. The vagina is short (derived from the urogenital sinus only) and ends blindly. The uterus and tubes are absent. The testes are normally developed but abnormally positioned. Testosterone production is normal or slightly increased. There is no problem of sex assignment because there is no trace of androgen activity. The diagnosis is likely when an individual presents following breast development at puberty, with primary amenorrhea, absent pubic and axillary hair, a short vagina, and an absent cervix and uterus. Males with enzyme defects that prevent testosterone synthesis will have a female phenotype, but breast development does not occur. Individuals with complete androgen insensitivity have augmented breast development because of a total absence of androgen influence. Androgen insensitivity accounts for about 10% of all cases of primary amenorrhea, third most common after gonadal dysgenesis and congenital absence of the vagina. The hormone profile in these individuals is typical: high LH, normal to slightly elevated male testosterone levels, high estradiol (for men), and normal to elevated FSH. Individuals with complete androgen insensitivity perform well in tests of visual-spatial ability, suggesting that androgens exert an organizing effect in the brain during development.

Incomplete Androgen Insensitivity

A spectrum of disorders, all due to an X-linked recessive trait, are known as incomplete forms of testicular feminization. Incomplete androgen insensitivity is one-tenth as common as the complete syndrome. The clinical presentation ranges from almost complete failure of virilization to essentially complete phenotypic masculinization. Between these poles exist examples of mild citoromegaly and slight labial fusion to significant genital ambiguity. Reifenstein's syndrome is now applied to all the intermediate forms that were initially given individual names (such as Lubs syndrome). Males have been described whose only indication of androgen insensitivity was azospermic or severe oligospermic infertility. Indeed, the incidence may approach 40% or more of men with infertility due to azospermia or severe oligospermia. However, the defect in androgen receptor function may be so subtle that some affected men are fertile. The undervirilized fertile male syndrome is another manifestation of this androgen receptor disorder. The diversity of presentation represents variable manifestations of the same mutant gene. The biochemical abnormality depends upon the degree of function of the androgen receptor or postreceptor events.

Molecular analysis of the androgen receptor gene in individuals with androgen insensitivity has demonstrated a spectrum of disorders in which both the complete and partial forms result from androgen receptor gene mutations. The gene encoding the androgen receptor is localized to the q11–12 region of the X chromosome. About 200 unique mutations have been identified.

Two types of defective androgen receptor function are recognized: abnormalities of androgen binding and abnormalities of DNA binding. The molecular defects responsible for these deficiencies have been identified and characterized. These include major structural abnormalities of the androgen receptor gene in which complete deletion of the gene or deletions of the exons encoding the androgen binding domain each result in the clinical picture of complete androgen insensitivity. In addition, point mutations that result in a defective receptor or alter receptor mRNA and cause reduced receptor protein production also result in complete androgen insensitivity. On the other hand, single base mutations that change a single amino acid yield subjects displaying either complete or partial androgen insensitivity. Alterations in receptor function, therefore, range from complete loss to subtle qualitative changes in the stimulation and transcription of androgen dependent target genes. Less understandable, however, is the poor correlation between receptor levels (and androgen binding affinity) with the degree of masculinization seen in partial androgen insensitivity. Nevertheless, the same mode of inheritance, despite differences in androgen receptor functioning, indicates that all forms originate in changes in the structural gene responsible for the androgen receptor.

Some men with partial androgen insensitivity can respond to androgen treatment with enhanced virilization. For example, a man with Reifenstein syndrome due to a mutation in the DNA-binding domain of the androgen receptor administration with high dose testosterone aromatization response with normal virilization. However, there is no test or measurement that will predict response, and only a few of these patients are able to respond to androgen therapy.

Molecular analysis of the androgen receptor gene in individuals with androgen insensitivity has demonstrated a spectrum of disorders in which both the complete and partial forms result from androgen receptor gene mutations. The gene encoding the androgen receptor is localized to the q11–12 region of the X chromosome.

The Androgen Insensitivity Syndromes

<table>
<thead>
<tr>
<th>Inheritance</th>
<th>5α-reductase</th>
<th>Complete</th>
<th>Incomplete</th>
<th>Reifenstein</th>
<th>Infertile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal</td>
<td>recessive</td>
<td>X-linked</td>
<td>recessive</td>
<td>X-linked</td>
<td>recessive</td>
</tr>
</tbody>
</table>

An absent uterus in a normally appearing female is encountered in only two conditions: androgen insensitivity and müllerian agenesis (the Mayer-Rokitansky-Kuster-Hauser syndrome). The latter is easily diagnosed because of the presence of pubic and axillary hair and a normal 46,XX karyotype (discussed in Chapter 11).

The “complete” form indicates that there is no androgen response; therefore, normal external female development occurs, and these infants should be reared as females. The testes (azoospermic with hyperplastic Leydig cells) may be present in the inguinal canals. Children with inguinal hernias and/or inguinal masses should be suspected of testicular feminization. There is virilization at puberty because of the lack of androgen response. In contrast to dysgenetic gonads with a Y chromosome, the occurrence of gonadal tumors is relatively late, rarely before age 25, and the overall incidence is less, about 5–10%. Therefore, gonadectomy should be performed at approximately age 16–18, to allow endogenous hormonal changes and a smooth transition through puberty.

Because of the importance of prophylactic gonadectomy, detection of this syndrome demands careful investigation for other affected family members. This syndrome follows an X-linked recessive pattern of inheritance. Apparent sisters of affected individuals have a 1 in 3 chance of being XY. Female offspring of a normal sister of an affected male have a 1 in 3 chance of being XXY. Female offspring of an affected male have a 1 in 6 chance of being XY. About a third of the patients have negative family histories and presumably represent new mutations.

In the past, conventional wisdom warned against unthinking and “needless” disclosure of the gonadal and chromosomal sex to a patient with complete androgen insensitivity. This attitude has changed as, more and more, patients desire and appreciate a full understanding of themselves. Although infertile, these patients are certainly completely female in their gender identity, and this should be reinforced rather than challenged. We now strongly advocate combining a truth telling education with appropriate psychological counseling of patient and parents. A good resource is the international support group, A + PS, that is headquartered in the United Kingdom (2 Shirburn Avenue, Mansfield, Nottinghamshire, NG18 2BY, UK). http://www.medhelp.org/www/aisc
5a-Reductase Deficiency

This form of familial incomplete male (46,XY) pseudohermaphroditism is due to an autosomal recessive trait that leads to a deficiency of the 5a-reductase enzyme (and, in some individuals, enzyme that is present but unstable) and is characterized by severe perineal hypoplasia and underdevelopment of the vagina. In the past it was known as pseudovaginal perineoscrotal hypoplasias (PPH). It differs from the incomplete forms of testicular feminization because, at puberty, masculinization occurs (the breasts remain male). Normal testicular function occurs, and there is no lack of response to endogenous or exogenous androgen. At birth, however, the external genitalia are similar to that of incomplete androgen insensitivity; i.e., hypoplasias, varying failure of fusion of labioscrotal folds and a urogenital opening, or separate urethral and vaginal openings. The cleft in the scrotum appears to be a vagina (there are no Müllerian ducts), and these patients have been reared as girls with an enlarged clitoris. At birth, steroid levels are normal, ruling out adrenal disorders.

Diagnosis can be established by demonstrating an elevated T:DHT ratio based upon the blood levels of testosterone and dihydrotestosterone, especially after HCG stimulation. The karyotype is XY, and, as with other incompletely masculinized males, the sex assignment is female if the phallus is inadequate. Gonadectomy is necessary to avoid not only neoplasia but also the virilization that is certain to appear at puberty. On the other hand, early correction of cryptorchidism and hypoplasias can preserve fertility and allow a male life. The deficiency is believed to be due to the homozygous state, manifest clinically only in males. Homozygous 46,XX females have reduced body hair, and although menarche may be delayed, fertility is normal.

At least 3 "types" of enzyme deficiency have been described in affected families:

1. Abnormally low concentration of enzyme.
2. Reduced enzyme activity due to enzyme instability.
3. Normal enzyme concentration but defective affinity for testosterone and/or essential cofactors leading to reduced enzyme activity.

Two 5a-reductase genes have been cloned. One isozyme (5a-reductase-1) is encoded on chromosome 5; mutations in the other isozyme (5a-reductase-2) encoded on chromosome 2 are responsible for male pseudohermaphroditism due to 5a-reductase deficiency. Because this deficiency in females does not affect fertility, this contributes to the high incidence of the disorder. The relatively easy switch of individuals reared as girls to boys at puberty suggests that the other 5a-reductase gene is operative in the brain.

Study of this syndrome points out important lessons in intersexuality. In this condition, the wolfian ducts virilize in a normal male fashion, but the urogenital sinus and genital tubercle persist as female structures. The failure is due to inadequate DHT formation intracellularly in these external genitalia tissues at the time the normal pentamic oocytes form. In the 5a-reductase deficiencies, the seminiferous vesicles, ejaculatory ducts, epididymis, and vas deferens, which are all testosterone-dependent, are present, whereas the DHT-dependent structures, external genitalia, urethra, and prostate, do not develop along male lines. Affected men have less facial and body hair, less temporal hairline recession, and no problems with acne. However, spermatogenesis, muscle mass, male libido, and deepening of the voice do occur in these men. DHT presence is a requirement only in the fetus, as indicated by the significant genital virilization these patients undergo at puberty and thereafter. These individuals require surgical correction of hypoplasias and cryptorchidism. Whereas the conversion from male to female role is exceedingly traumatic psychologically, the reversal of sex identity (female to male) of some of these patients have undergone at puberty was apparently uncomplicated.

In one such case, a "double-life" was conducted. Although functioning in all public respects as a female, one 5a-reductase individual conducted numerous and prolonged heterosexual affairs, which were quite satisfactory, albeit clandestine. He had known of his male sexual identity since puberty but delayed medical assistance for fear that exposure would bring shame and guilt to his religiously devoted elder "old world" mother. He decided to keep his secret until his mother died. He finally sought diagnostic help at age 65, however, because his mother at age 93 continued to enjoy good health.

Abnormal Androgen Synthesis

Defective male development may stem from a secretory failure of the testes during the critical period of sex differentiation. In addition to the obvious specific and often familial defects in enzymatic steps leading to testosterone biosynthesis, a variety of other intrinsically testicular problems can lead to male pseudohermaphroditism. In all the following conditions account for 4% of male pseudohermaphroditism:

1. Aberrations in testicular organogenesis (dysgenetic testes).
2. Gonadotropin resistant testes (LH receptor mutations).
4. Defective synthesis, secretion, or response to anti-müllerian hormone.

Defects in testosterone synthesis can be at any one of the 3 required enzymatic reactions that lead from cholesterol to testosterone: 3b-hydroxysteroid dehydrogenase, P450c17, and 17b-hydroxysteroid dehydrogenase. These defects are inherited as autosomal recessive traits, and the phenotypes range from partial to complete male pseudohermaphroditism.

Patients with male pseudohermaphroditism who are considered variants of testicular feminization upon partial virilization at puberty may actually have a defect in androgen synthesis. The diagnosis is made by demonstrating elevated blood levels of androstenedione and estrogens, while the blood level of testosterone is low or normal. When the enzyme involves a reaction that is active in the adrenal gland (all but the 17b-hydroxysteroid dehydrogenase), the adenals blocks are usually severe with adrenal failure and death in the newborn period.

The male pseudohermaphroditism due to deficient testicular 17b-hydroxysteroid dehydrogenase activity because of mutations in the encoding gene has male internal genitalia and no Müllerian structures. The characteristic clinical findings in these patients are external female genitalia at birth with testes usually located in the inguinal canal. Raised as females, these individuals undergo virilization at puberty because of an increase in testosterone levels. This increase in pubertal testosterone is derived by conversion of elevated circulating levels of androstenedione (in response to the increase in LH at puberty) by unaffected 17b-hydroxysteroid dehydrogenase isoenzymes. In individuals being raised as girls, early gonadectomy is required to avoid virilization at puberty and testicular neoplasia.

Gonadotropin Resistant Testes

Male pseudohermaphrodites, due to agenesis or abnormal differentiation of Leydig cells, are characterized by reduced responsiveness to LH/HCG. All these cases can be formed "gonadotropin resistant testes." In general, the characteristics of the syndrome include basically female but ambiguous genitalia, male cryptorchid testes with degenerated Leydig cells (Leydig cell hypoplasia or agenesis), no Müllerian ducts but present vas deferens and epididymis, elevated gonadotropins (FSH rises further after gonadectomy, indicating the presence of inhibin).

Molecular studies have revealed that these XY gonadotropin resistant individuals with varying degrees of masculinization (from total feminization to nearly normal male genitalia) reflect the amount of Leydig cell function allowed by mutations in the LH receptor. These mutations are inherited by recessive transmission, and, therefore, heterozygous parents have no characteristic affected phenotypes.

Congenital Lipoid Adrenal Hyperplasia

Most of the cholesterol used for steroid synthesis is derived from the mobilization and transport of intracellular stores. Indeed, the rate-limiting step in

<table>
<thead>
<tr>
<th>Spermatogenesis</th>
<th>Decreased</th>
<th>Absent</th>
<th>Absent</th>
<th>Absent</th>
<th>Absent</th>
<th>Decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Müllerian</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>Wolfian</td>
<td>Male</td>
<td>Female</td>
<td>Female</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>External</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
</tr>
</tbody>
</table>

Breasts: Male, Female. Gynecomastia: Male, Female.
Almost all (99%) of conceptuses with only one X chromosome abort. The remaining 1% account for an incidence of Turner syndrome between 1 in 2500 and 1 in 5000 that remain throughout life.

Syndrome have normal intelligence; however, there may be difficulty with mathematical ability, visual-motor coordination, and spatial-temporal processing, problems condition. It is important to assess the aorta, aortic root, and aortic valve with ultrasonography at least in infancy and again during the teens. Patients with Turner and lack of sexual development become apparent. At birth, however, lymphedema (due to hypoplasia of superficial vessels) may indicate the presence of abnormal gonia
deviation or mosaicism with an abnormal X. Henry Hughbert Turner, born in 1892, became chief of endocrinology and associate dean of the University of Oklahoma school of medicine. He was president of the Endocrine Society in 1968, and died in 1970. Turner's clinical description of this syndrome was presented to the medical community in 1931.

Almost 60% of Turner patients have the total loss of one X chromosome; the remainder have either a structural abnormality in one of the X chromosomes or mosaicism with an abnormal X. Gonadal dysgenesis with bilateral rudimentary streak gonads due to an abnormality in or absence of one of the X chromosomes in all cell lines is called Turner syndrome. Approximately 60% of Turner patients have the total loss of one X chromosome; the remainder have either a structural abnormality in one of the X chromosomes or mosaicism with an abnormal X. Henry Hughbert Turner, born in 1892, became chief of endocrinology and associate dean of the University of Oklahoma school of medicine. He was president of the Endocrine Society in 1968, and died in 1970. Turner's clinical description of this syndrome was presented to the medical community in 1931.

Gonadal dysgenesis 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 a protein activation stimulated by cholesterol hormones. Several proteins have been characterized and proposed as regulators of acute intracellular cholesterol transfer. Steryl carrier protein 2 (SCP2) is able to bind and transfer cholesterol between compartments within a cell. Another candidate is a small molecule, steriodogenesis activator polypeptide (SAP), and still another is potential protein receptor (PSR), which affects cholesterol flux through a pore structure. But the most studied and favored protein as a regulator of acute cholesterol transfer is steriodogenesis acute regulator (SAR) protein. SAR messenger RNA and protein are induced concomitantly with acute steriodogenesis in response to cyclic AMP stimulation; SAR increases steroid production; and SAR is imported and localized in the mitochondria. But most impressively, congenital adrenal hyperplasia (an autosomal recessive disorder) has been demonstrated to be a failure in adrenal and gonadal steriodogenesis due to a mutation in the SAR gene, which results in premature stop codons. With this mutation, a low level of steriodogenesis is possible, even permitting feminization at puberty, but continuing trophic hormonal stimulation results in an accumulation of intracellular lipid deposits that destroy steriodogenic capability.

SAR mediates the transport of cholesterol into mitochondria in adrenal and gonadal steriodogenesis, but not in the placenta and brain. It is synthesized in a precursor form as a 285-amino acid protein that has a 25-residue sequence cleaved from the NH2-terminal protein after transport into mitochondria. The mutant forms of SAR undergo premature truncation that prevents this proteolytic cleavage. Mutations of the SAR gene, located on chromosome 8p11.2, are the only inherited disorder of steriodogenesis not caused by a defect in one of the steriodogenic enzymes. The absence of SAR expression in placenta and brain indicates the presence of different mechanisms for cholesterol transport in those tissues.

Abnormal Anti-Müllerian Hormone

Hernia Uterine Inguinale (Uterine Hernia Syndrome)

Individuals with this syndrome appear to be normal males, but relatively well-differentiated müllerian duct structures are found, usually a uterus and tubes in an inguinal hernia sac. This is due to a failure of AMH function either as a result of failure of Sertoli cell secretion of this polypeptide or an inability of the müllerian ducts to respond to AMH. Therefore, this syndrome is due to mutations in either the gene for AMH or its receptor. The mutation is inherited as a recessive trait, either X-linked or autosomal. Fertility is usually preserved.

Disorders of Gonadal Development

The proper development and eventual function of the gonad depend on the presence of germ cells, the appropriate sex chromosome constitution, and appropriate gonadal ridge somatic cells. Errors in meiotic division can cause aneuploidy and abnormal sex chromosomes. These occur by nondisjunction, anaphase lag, translation, breakage, rearrangements, or deletions. Mitosis can also be marred by nondisjunction and anaphase lag leading to mosaicism. Two or more different cell lines can persist and appear in different tissues. Finally, abnormal gonadogenesis may occur as a result of structural or disease related catastrophes leading to loss of fetal gonadal function.

Bilateral Dysgenesis of the Testes (Swyer Syndrome)

Affecting individuals have an XY karyotype but normal (infantile) female external and internal genitalia. There are fibrous bands in place of the gonads yielding primary amenorrhea and lack of secondary sexual development at puberty. It is a matter of prudent practice to avoid the possibility of virilization or neoplasms; therefore, removal of these bands is advocated as soon as the diagnosis is made. Presumably, tests fail to develop or were eliminated (testicular regression) before internal or external genital differentiation. At least one cause of this syndrome is a mutation in the SRY gene. Estrogen and progestin sequential therapy supports female secondary sex development.

Anorchia

Affecting XY individuals have infantile unambiguous male external genitalia, and male Wolffian ducts and lack müllerian ducts. There are, however, no detectable testes. Early testsis function did occur (wolffian presence, AMH function) but was not sustained in sufficient amounts or duration to develop a normal size phallus. It is frequently called “the disappearing testsis syndrome.” Sex of assignment depends on the extent of external genitalia development.

True Hermaphroditism

Hermaphroditus, the Greek god with bisexual attributes, was the son-daughter of Hermes, the god of athletics, secrets, and occult philosophy, and Aphrodite, the goddess of love. The bisexual theme was immortalized in countless statues by the Greeks and Romans, depicting a normal woman with normal male external genitalia (not a combination commonly encountered in real life). Piny (23–79 AD) was the first to apply the term hermaphrodithe to humans, presenting a description in his massive work, Historia Naturalis.

Abnormal sexual differentiation can occur as a result of a mixture of gonadal sex (true hermaphroditism) or complete uncertainty of gonadal sex (gonadal dysgenesis with some virilization). A true hermaphrodit possesses both ovarian and testicular tissue. Both types may be contained in one gonad (ovotestis) or less often, one side may be an ovary, the other a testis. The internal structures correspond to the adjacent gonad. In the majority, external genitalia are ambiguous with sufficient male character to allow male sex assignment. However, three-fourths develop gynaecomastia and half menstruate after puberty. Sixty percent are genetic females (XX), few are XY, the rest are mosaics with at least one cell line XX. 46,XX individuals without SRY may have a mutation of an autosomal gene that permits testicular determination in the absence of the tests determining factor.

Gonadal Dysgenesis

Gonadal dysgenesis with bilateral rudimentary streak gonads due to an abnormality in or absence of one of the X chromosomes in all cell lines is called Turner syndrome. Approximately 60% of Turner patients have the total loss of one X chromosome; the remainder have either a structural abnormality in one of the X chromosomes or mosaicism with an abnormal X. Henry Hughbert Turner, born in 1892, became chief of endocrinology and associate dean of the University of Oklahoma school of medicine. He was president of the Endocrine Society in 1968, and died in 1970. Turner’s clinical description of this syndrome was presented to the annual meeting of the Association for the Study of Internal Secretions in 1933.

In the absence of gonadal development, these individuals are phenotypic females. The well-known characteristics are short stature (142–147 cm, 56–58 inches), sexual infantilism, and streak gonads. The streak gland is composed of white fibrous stromal tissue, 2–3 cm long and about 0.5 cm wide, containing no vasa or follicular derivatives. Other congenital problems in this syndrome are a webbed neck, a high arched palate, cubitus valgus, a broad shield-like chest with widely spaced nipples, a low hairline on the neck, short fourth metacarpal bones, disproportionately short legs, and renal abnormalities (horseshoe kidney, unilateral pelvic kidney, rotational abnormalities, and partial or complete duplication of the collecting system). Autoimmune disorders are common, such as Hashimoto's thyroiditis, Addison's disease, alopecia, and vitiligo. Hypothyroidism is present in about 10% of young patients and may reach 50% later in life. Mild insulin resistance and hearing loss are also common. Patients with Turner syndrome are very prone to keloid formation. One-third of patients with Turner syndrome have cardiovascular abnormalities, including bicuspid aortic valves, coartation of the aorta, mitral valve prolapse, and aortic aneurysms. Often the diagnosis is not made until puberty when amenorrhea and lack of sexual development become apparent. At birth, however, lymphedema (due to hypoplasia of superficial vessels) of the extremities may indicate the condition. It is important to assess the aorta, aortic root, and aortic valve with ultrasonography at least in infancy and again during the teens. Patients with Turner syndrome have normal intelligence; however, there may be difficulty with mathematical ability, visual-motor coordination, and spatial-temporal processing, problems that remain throughout life.

Almost all (99%) of conceptuses with only one X chromosome abort. The remaining 1% account for an incidence of Turner syndrome between 1 in 2500 and 1 in 5000
Because of the high incidence of assorted abnormalities in patients with Turner syndrome, the following evaluations should be performed, some only once at the time of diagnosis, and others annually as part of on-going surveillance: thyroid function testing (annually) and antibodies (at least once), intravenous pyelogram or renal ultrasonography (once if normal), echocardiography, audiometry, and annual evaluation of the lipid profile and glucose metabolism.

Spontaneous puberty has been reported in 10–20% and spontaneous menstruation in 2–5% of girls with Turner syndrome. Rare pregnancies have been reported in those who have had spontaneous menstruation. 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. When pregnancy does occur in an X deficient subject, the infant has a 30% chance of congenital anomalies, including Down syndrome, spina bifida, and congenital heart disease.

A large variety of mosaic patterns is seen with gonadal dysgenesis. From analysis of the various combinations, it is apparent that short stature is related to loss of regions on the short arm of one X chromosome. Distal arm deletions of one of the X chromosomes are associated with amenoaroea (usually secondary after some ovarian function) and streak gonads, but the patients are not always growth compromised nor do they display other Turner somatic malformations. Long arm deletions near the centromere are associated with primary amenorrhea. Thus, loss of material from the short arms of the X chromosome leads to short stature and the other stigmata of Turner syndrome. This suggests that normal ovarian development requires two loci, one on the long arm and one on the short arm; loss of either results in gonadal failure. Thyroid autonomy is common in Turner syndrome, but Hashimoto's thyroiditis may be specific to 46,XXq cases. It should be emphasized that any of the Turner syndrome characteristics can be seen with any karyotype.

Just as X chromosome monosomy, with deletion of the second X chromosome, results in Turner's phenotype, the same will apply to loss of the Y chromosome. The 45,X karyotype derived from leukocyte culture does not guarantee that a mosaic does not exist with a gonadal line containing XY. For this reason, annual pelvic examinations and appropriate screening are required to detect incipient signs of gonadal neoplasia as an adenocinar mass. If a presumed 45,X patient develops breasts or sexual hair without exogenous therapy, a gonadoblastoma or dysgerminoma should be considered and ruled out. Heterosexual signs require scrutiny in all 45,X individuals. Expert consultation should be obtained to pursue further analysis with X- and Y-specific DNA probes. Why not screen all patients with some form of gonadal dysgenesis with molecular techniques? For example, 3–4% of 45,X patients had Y-derived DNA detected by polymerase chain reaction analysis. In one series of 40 Turner syndrome patients, one patient had a useful Y chromosome material using polymerase chain reaction of the gene from the SRY region of the Y chromosome in 3 of 18 patients (none of whom had evidence of Y-chromosomal material by cytogenetic analysis). Y-chromosomal segments from the SRY region were detected by polymerase chain amplification and Southern blot analysis. Therefore the prevalence of undetected Y material is low. Furthermore, is it necessary to know of the presence of silent Y mosaicism?

The problem is that we do not know if there is any clinical significance to Y-derived material detected only by molecular techniques. Another shortcoming of this approach is that polymerase chain reaction and in situ hybridization identify only those DNA sequences that correspond to the probes used in the analysis. There may be a sequence from a Y chromosome that does not correspond to the selected probes. A totally reliable method awaits identification of the gene responsible for tumor formation or the complete mapping of the Y chromosome sequences. Until then, the recommendation is to seek Y chromosomal material in patients with virilization or when a chromosomal fragment of uncertain origin is identified on the karyotype.

The term “gonadal dysgenesis” is frequently used to describe all subjects with female genitalia, normal müllerian structures and streak gonads (with either 46,XX or 46,XY karyotypes). Not all possess the spectrum of Turner phenotype anomalies, and some show none of these characteristics. The latter groups are referred to as “pure gonadal dysgenesis.” Ovarian development and subsequent internal and external genitalia development proceed normally. However, at variable times thereafter, the gonads undergo accelerated germ cell loss, and premature degeneration of the ovaries ensues. A similar morphologic consequence of premature ovarian degeneration exists in 46,XX gonadal dysgenesis. In these instances, a mutation in an autosomal gene is considered the likely etiologic factor, inherited in autosomal recessive fashion. PERRTAURISM is the combination of XX gonadal dysgenesis and neurosensory deafness.

Subjects with 46,XY karyotype can also have gonadal dysgenesis; most, but not all, with the pure gonadal dysgenesis form are without Turner stigmata. The 46,XY gonadal dysgenesis patients are diagnosed in early adolescence with delayed pubertal development. As expected they show elevated gonadotropins, normal female levels of androgen, and low levels of estrogen, female external genitalia, a uterus, and fallopian tubes. Minimal breast enlargement reflects peripheral aromatization of androgens. Menstrual function suggests tumor development in the streak gonad. These streaks often display ovarian stroma but no follicles. Their propensity to tumor development is significant, a 20–30% incidence. Patients with mosaic patterns in the karyotype have a reduced risk of tumor, but it still amounts to 15–20%. The most common tumor is the often bilateral gonadoblastoma, but dysgerminomas and even the more threatening embryonal carcinoma are also seen. Intraabdominal testes should be removed as early in life as possible because of the known risk of tumor development. If well visualized, streak gonads can be removed by laparoscopy, and the uterus and tubes should be retained for the possibility of pregnancy with donor oocytes.

The etiology of this defect is thought to be a short arm Y chromosome deletion involving SRY, a mutation in other genes that leads to inhibition of SRY function, or a XXq mutation of SRY. There is evidence to suggest that some cases of 46,XY gonadal dysgenesis are due to impaired function of a gene on the X chromosome that is necessary for normal SRY function. The absence of Turner stigmata in these patients suggests the preservation of a nearby gene that protects XY individuals from Turner development. In 46,XY partial gonadal dysgenesis individuals there is some testicular development, and therefore, they present as newborns with ambiguous genitalia. The degree of external masculinization and the relative proportions of müllerian and wolffian duct structures present correlate with the extent of testes differentiation.

Curiously, to date analysis of a large number of 46,XY gonadal dysgenesis patients has not demonstrated mutations of the coding region of the SRY gene. The association of this form of gonadal dysgenesis with duplications of the short arm of the X chromosome, and various inherited syndromes of multiple congenital anomalies, indicate the role of several genes other than SRY in testes development. There are several reports of 46,XY male pseudohermaphroditism because of mutations in the LH receptor gene. The 46,XX sisters of these individuals present with primary amenorrhea and a normal female phenotype.

Males who have a 46,XX karyotype usually have XXY present, probably by translocation from the Y to an X chromosome. Some subjects, however, lack SRY, indicating that more than one genetic defect yields 46,XX maleness.

Multiple X females (47,XXX) have normal development and reproductive function, although mental retardation may be more frequent. Secondary amenorrhea and/or eunuchoidism can be seen.

### Mixed Gonadal Dysgenesis

Mosaic involving the Y chromosome can be associated with abnormalities of sex differentiation. Of the variety of karyotypes possible from loss of the Y by nondisjunction, 45,X/46,XY is the most common. A wide variety of phenotypes is displayed by these individuals from newborns with ambiguous genitalia to normal fertile males or normal female phenotype with bilateral streak gonads. Most have short stature, and one-third have other Turner stigmata.

In the “typical” mixed gonadal dysgenesis case presenting with abnormal sex differentiation, the usual gonadal pattern is a streak gonad on one side and a dysgenetic or normal appearing testis on the other side of the abdomen. Müllerian and wolffian duct development correlates with the character of the ilipaternal gonad. All possible permutations combining streaks, and dysgenetic and apparently normal testes have been encountered. The diversity of presentation is presumed to reflect...
Surgical Removal of Gonadal Tissue

There is no debate that gonadal tissue having any Y chromosome component in phenotypic females requires removal as soon as the diagnosis is made to avoid the risk of malignant gonadal tumors. There is one exception to this rule. Because gonadal tumors occur relatively late in patients with complete androgen insensitivity, surgery is delayed until after puberty. An accomplished laparoscopic technique can attempt this procedure, with the option of laparotomy if the gonads prove to be inaccessible. Streak gonads have been removed in this fashion, as well as the testes in androgen insensitivity. With androgen insensitivity, the gonads can be close to the external iliac artery and herniated into the inguinal canal if the procedure is more difficult, and care must be taken to extract the gonad from the inguinal canal to prevent complete excision. It may also be necessary to make a small abdominal incision or a culdotomy to extract the gonad in order to avoid morcellation. Preoperative magnetic resonance imaging can be helpful in order to accurately localize gonadal tissue and select the correct operative approach.

The uterus and tubes should be preserved for the possibility of pregnancy with donor oocytes.

Hormone Treatment of Patients Without Ovaries

When ovaries are absent in individuals being reared as females, either because of surgery or streak gonads, hormonal treatment will be necessary at puberty and thereafter. Estrogen will initiate and sustain maturation and function of secondary sexual characteristics, and promote the achievement of the full height potential. The adolescent increase in bone density is a very important determinant of an individual’s later risk for osteoporosis. This alone is sufficient reason for treatment. Very small amounts of estrogen will promote growth and development. Early treatment with estrogen, however, can reduce overall growth. Start at about age 12–14 with unopposed estrogen (0.3 mg conjugated estrogens or 0.5 mg estradiol daily). After 6 months to 1 year, move to a sequential program with 0.625 mg conjugated estrogens or 1.0 mg estradiol daily and 5 mg medroxyprogesterone acetate or an equivalent progestin for the first 14 days each month (if a uterus is present). Although not proven by appropriate studies, most clinicians believe an increase in estrogen dose to 1.25 mg conjugated estrogens or 2.0 mg estradiol is warranted (along with adequate calcium intake) to maximize bone and breast responses. Adequacy of treatment can be assessed by following bone age changes, although this is unnecessary in most cases. In patients with genetic short stature in stature (e.g., Turner syndrome), estrogen treatment is not started until bone age is greater than 12 to avoid epiphiyal closure and to allow a longer period of time for bone growth. Clinicians should strive for good compliance with lifelong hormone therapy in order to protect against osteoporosis and cardiovascular disease. Patients can receive helpful education and support from the Turner’s Syndrome Society of the United States, 1313 SE Fifth St., Suite 327, Minneapolis, MN 55414, http://www.turner-syndrome-us.org/

Stimulation of Growth

Short stature occurs in virtually all patients with a 46,XY karyotype and nearly all patients with Turner syndrome who have other karyotypes. This growth impairment begins in utero, is apparent throughout childhood, and results in a short adult height (a mean of 143 cm, about 4 feet 8 inches). This attainment of growth is partly due to insufficient growth hormone secretion because of the deficiency in sex steroids and also to an end organ resistance to insulin-like growth factor-I. Anabolic steroids have been used to stimulate growth, especially in patients with Turner syndrome. Short-term growth can be stimulated by anabolic steroids; however, the effect on final adult height is equivocal because epiphyseal maturation is also enhanced. Furthermore, virilizing side effects are a drawback. The combination of low doses of an anabolic steroid (e.g., oxandrinolone) and recombinant growth hormone under the direction of a pediatric endocrinologist offers the best prospect. Treatment with growth hormone (50 µg/kg/day) yields significant growth acceleration that can be sustained for at least 6 years, achieving an adult height as much as 10 cm over the initial predicted height, and most patients can achieve a height that is greater than 150 cm (about 4 feet 11 inches). The future may see effective use of growth hormone-releasing hormone for this purpose. It is worth noting that adolescents with Turner syndrome either treated with growth hormone or not do not lose bone mineral density prior to estrogen treatment; hence, it is unnecessary to begin estrogen treatment at an early age (when it might counteract the goal of growth hormone therapy: achieving maximal height).

Now that the success of growth hormone treatment is recognized and accepted, an argument can be made for chromosomal screening by molecular analysis of all growth-retarded girls. In a screening of 375 mildly growth-retarded girls, 18 cases of Turner syndrome were identified; 14 of whom had none of the typical clinical features.

The Possibility of Pregnancy

In women who have variants of gonadal dysgenesis and who menstruate, pregnancy can occur. However, there is a 30% incidence of congenital anomalies in the offspring, including spina bifida and Down syndrome. Sex chromosome abnormalities are frequent in the children born to mosaic mothers. Prenatal diagnosis by amniocentesis or chorionic villus biopsy is highly advised. Assisted reproductive technology with donated oocytes yields excellent results in women with streak gonads (see Chapter 31). There is an important consideration in regards to those patients who become pregnant. Fatal aortic events (aneurysm, dissection, or rupture) can occur during pregnancy in patients with gonadal dysgenesis. A cardiology consultation with an echocardiogram is strongly advised for these women prior to pregnancy.

Noonan Syndrome

Both affected males and females with Noonan syndrome have apparently normal chromosome complements and normal gonadal function. The phenotypic appearance of the female is that of a patient with Turner syndrome: short stature, webbed neck, shield chest, and cardiac malformations. The cardiac lesions, however, are different. Pulmonic stenosis is most frequent in Noonan syndrome as opposed to aortic coarctation in Turner syndrome. In the past these patients have been referred to as male Turner’s or Turner’s with normal chromosomes. Noonan’s are fertile and transmit the trait as an autosomal dominant with variable expression.

Diagnosis of Ambiguous Genitalia

Ambiguous external genitalia in a newborn infant represents not only a major diagnostic challenge, but a social and medical emergency. The physician is involved in a pressure-filled situation because of the necessity for making such an influential decision as the sex of sexual rearing. Rapid and organized evaluation must be initiated to assign the appropriate gender, identify a possible life-threatening medical condition, and begin necessary medical, surgical, and psychological interventions. Input from a team of experts in endocrinology, genetics, neonatology, psychology, surgery, and urology is essential. Nevertheless, the primary clinician should be the sole contact with the family. Diagnostic procedures may delay the decision, but it is well recognized that a period of delay is far better than later reversal of the sex assignment. Naming of the child should be delayed until a gender is firmly assigned. Parental education and guidance are essential in this anxiety-ridden situation.

The most important point to remember when confronted with a newborn infant with ambiguous genitalia, or an apparently male infant with bilateral cryptorchidism, is that the prime diagnosis until ruled out is congenital adrenal hyperplasia. The reason is clear: adrenal hyperplasia is the only condition that is life threatening. Signs of adrenal failure such as vomiting, diarrhea, dehydration, and shock may develop rapidly. Furthermore, most infants with ambiguous genitalia are virilized females, and most of these have congenital adrenal hyperplasia.

The history of a previously affected relative may aid in the diagnosis of testicular feminization or any of its variants. Similarly, the history of a sibling with genital ambiguity or the history of a previous neonatal death in a sibling strongly suggest the possibility of adrenal hyperplasia. A history of maternal exposure to androgenic compounds may be difficult to elicit. The mother may be unaware of the nature of her medications, and the obstetrician should be consulted to determine if medication was used for threatened or recurrent miscarriage or endometriosis. Although the appearance of the external genitalia in intersex infants may be similar regardless of etiology, and a definitive diagnosis unachievable by physical examination alone, certain useful clues can be discerned.

Are gonads palpable? Palpation of the genital and inguinal regions is the most important part of the physical examination. Gonads in the inguinal regions or in scrotal folds are almost certainly testes. Ovaries are not found in scrotal folds or in the inguinal regions. The testicles, however, may be intraabdominal. If testicles are
not palpable, the infant should be considered to have congenital adrenal hyperplasia until demonstrated otherwise.

**What is the phallus length and diameter?** Measured from the pubic ramus to the tip of the glans, a stretched penile length of less than 2.5 cm is 2.5 standard deviations below the mean for infants at 40 weeks of gestational age (2.0 cm for 36 weeks and 1.5 cm for 32 weeks). The normal newborn clitoris measures less than 1 cm long; a normal newborn penis measures 2.8–4.2 cm in length.

**What is the position of the urethral meatus?** The urethral meatus can range from a mild hypospadias to an opening in the perineal area into a urogenital sinus. Hypospadias is almost always accompanied by chordee, which is a ventral curvature of the phallus resulting from a shortened urethra.

**To what degree are the labioscrotal folds fused?** The findings can range from unfused labia majora of a normal female through labia with variable degrees of posterior fusion, a bifid scrotum, to a fully fused normal appearing male scrotum. The distance from the anus to the edge of the vagina divided by the distance from the anus to the base of the clitoris is a ratio which is less than 0.5 in normal females. A ratio greater than 0.5 indicates some degree of labioscrotal fusion.

**Is there a vagina, vaginal pouch, or urogenital sinus?** Does the rectal exam suggest a midline structure that might be a uterus? A uterus can be palpable, especially shortly after birth when the uterus is a little enlarged in response to maternal estrogen.

Further important physical signs include evidence of hyperpigmentation due to the melanocyte stimulation associated with high levels of ACTH in adrenal hyperplasia, dehydration, hypotension, hypertension, and the manifestations of Turner syndrome such as webbed neck, low hairline, edema of hands and feet, and cardiac and renal anomalies. Careful examination of the phallus may differentiate between a clitoris and a penis. The penis has a midline ventral frenulum, while the clitoris has two folds which extend from the lateral aspects of the clitoris to the labia minora.

All patients with ambiguous genitalia require pelvic ultrasonography (to detect a uterus and ovaries or undescended testes) and a retrograde injection of contrast media into the urogenital orifice to outline the urethra and/or vaginal anatomy, the existence of a urogenital sinus, and the presence of a cervix. Magnetic resonance imaging of the newborn pelvis can supplant both of these procedures.

Rapid testing of blood leukocytes for karyotype analysis, serum electrolytes, serum androgens (androstenedione, testosterone, dehydroepiandrosterone, dehydroepiandrosterone sulfate), 17-hydroxyprogesterone, 11-deoxycorticosterone, and 11-deoxycortisol is an essential laboratory component in the evaluation of intersex. In selected circumstances, ACTH testing and genital skin fibroblasts provide specific amplifying information. Currently, there is no rapid test using DNA probes for diagnosing the etiology of ambiguous genitalia.

**Differential Diagnosis**

The presence or absence of palpable gonads, the presence or absence of a uterus, and the karyotype places the patient in one of four categories: female pseudohermaphroditism, male pseudohermaphroditism, true hermaphroditism, or gonadal dysgenesis.

**Clinical signs of adrenal failure indicate that the newborn has some form of adrenal enzyme defect regardless of the steroid pattern. The diagnosis is certain if such an infant is hyperkalemic and hyponatremic (due to aldosterone deficiency) or hypertensive and hypokalemic (secondary to elevated deoxycorticosterone).**

In the absence of maternal androgen excess, the diagnosis of genetic females with excess androgen (female pseudohermaphroditism) must distinguish 3 forms of congenital virilizing adrenal hyperplasia. The diagnosis of 21-hydroxylase deficiency is confirmed by finding an elevated level of 17-hydroxyprogesterone in the serum. Elevated levels of 11-deoxycorticosterone and 11-deoxycortisol are found in 11b-hydroxylase deficiency, while the precursors to cortisol and androgen, 17-hydroxypregnenolone and dehydroepiandrosterone, are high in the 3b-hydroxysteroid dehydrogenase deficiency. Marginal basal aberrations can be accentuated by an ACTH stimulation test.

Male pseudohermaphroditism (genetic males with too little androgen) can be the result of one of the relatively rare enzyme disorders. Four derive from enzymatic errors in the biosynthesis of testosterone. P450c17 deficient patients display elevated progesterone, whereas individuals with 3b-hydroxysteroid dehydrogenase deficiency have elevated DHA and 17-hydroxypregnenolone levels. Again, ACTH testing can be useful in amplifying compensated defects. A testosterone biosynthetic error which leads to deficient masculinization involves an enzyme not required for glucocorticoid or mineralocorticoid synthesis. Thus no evidence of adrenal insufficiency with electrolyte disturbance is seen in 17b-hydroxysteroid dehydrogenase deficiency. Using HCG stimulation testing, elevated androstenedione and DHA are found in congenital virilizing adrenal hyperplasia. The diagnosis of 21-hydroxylase deficiency is confirmed by finding an elevated level of 17-hydroxyprogesterone in the newborn. An elevated ratio of testosterone to dihydrotestosterone after HCG stimulation suggests 5a-reductase deficiency that can be confirmed in culture. On the other hand, patients with androgen insensitivity will have normal ACTH and HCG stimulation tests and abnormal androgen binding or androgen receptor gene mutations in cultured cells.

Although laparotomy is not necessary for assignment of sex, it may be the only way to arrive at a definitive diagnosis. Laparotomy is indicated in the following situations (laparoscopic evaluation is inadequate because gonads may be small and hidden in the inguinal canal).

1. The XX infant with ambiguous genitalia, normal androgens, in apparent good health, and no history of maternal androgen exposure. This is either a true hermaphrodite or a variant of mixed gonadal dysgenesis, and gonadectomy is indicated.
2. The XY patient with ambiguous genitalia, without palpable gonads, and normal androgens. The possibilities are incompletely masculinized males (variants of testicular feminization), a true hermaphrodite, mixed gonadal dysgenesis, and 5a-reductase deficiency. Sex of rearing will be female, and gonadectomy is necessary to avoid virilization at puberty and the propensity to develop gonadal neoplasia.

Only laparotomy, gonadal biopsy, and/or gonadectomy confirm the diagnosis of true hermaphroditism in individuals with ambiguous genitalia and 46,XY or 45X/46,XY karyotypes. It should be emphasized that evaluations for mutations in the SRY gene are powerful research tools, but these tests are not currently available for clinical determinations.
Assignment of Sex of Rearing

In a newborn who presents a problem of correct sex assignment, it is better to delay than to reverse the sex assignment at a later date. Generally, the decision can be made within a few days, at most a few weeks. In dealing with the parents, terms with unfortunate connotations, such as hermaphroditism, should be avoided. An easy way to explain ambiguous genital development to parents is to indicate that the genitals are unfinished, rather than abnormal from a sexual point of view.

Despite the wealth of knowledge derived from the world of molecular biology, the decision regarding sex of rearing remains a clinical judgment. When all the information is in place, gender assignment will rest on:

1. Future fertility.
2. The projected appearance of genitalia after puberty.
3. Penile adequacy for coital function.

The future fertility in all masculinized females is unaffected. With proper treatment, reproduction is possible, since the internal genitalia and gonads are those of a normal female. Therefore, all masculinized females should be reared as females.

The only other category of patients with ambiguous genitalia with reproductive capability consists of males with 1) isolated hypospadias, 2) the male with repaired isolated cryptorchidism, and 3) the male with the uterine hernia syndrome.

All other patients with ambiguous genitalia will be sterile. Except for salt-wasting adrenal hyperplasia, the physician's prime concern is not with physical survival, but to enable the patient to grow into a psychologically normal, healthy, and well-adjusted adult. The sex of assignment depends upon only one judgment: can the phallus ultimately develop into a penis adequate for intercourse. The success of a penis is dependent upon erectile tissue, and the genitalia should not only be serviceable but also erotically sensitive. Technically, the construction of female genitalia is easier, and therefore the physician must be convinced that a functional penis is possible.

All decisions regarding sex of rearing and the overall treatment program should be made early in life. If a case has been neglected, sex reassignments must be made according to the gender identity in which a child has developed. Reassignment of sex can probably be made safely up to age 18 months.

Socialization and hormone therapy are important for gender identity and sexual function. Future gender role and identity can be in accord with assigned sex if 4 conditions are met:

1. The patients are comfortable in their ability to raise their child and their resolution of any doubts or uncertainty about the sex of the child. In this acceptance and adaptation, the parents must have participated in and agreed to the sex assignment decision. Ambiguous names should be discouraged to ensure normal gender identity.
2. Genital reconstruction should take place as early as possible, certainly well before 18 months. Thereafter, sex reassignment is difficult and adjustment impaired.
3. Properly timed hormonal and/or additional surgical interventions must be provided at puberty.
4. The patient should be informed about his or her condition as deemed age-appropriate.

Sex Testing in Athletics

Gender verification tests were introduced into competitive sports in the 1960s because of the concern that individuals might masquerade as the opposite sex to gain a competitive advantage. Sex testing can be based on chromosomal or other histological assays, hormone measurements, or anatomical criteria. Such testing could identify potential gender-ambiguous subgroups of problem athletes: transsexuals who have undergone sex change operations, impersonators, and individuals with intersex conditions. Neither sex changed individuals nor masqueraders have been in a productive athletic events. However, sex testing creates the possibility of public exposure and embarrassment for intersex individuals. In addition, sex testing has the potential for revealing to an unaffected individual an underlying diagnosis in a shocking, difficult manner. For all of these reasons, it is argued that physical examination alone should suffice, and any individual with a female phenotype should be allowed to compete as a woman. Even individuals with mild disorders would be able to pass the physical examination requirement. Gender testing of athletes is unnecessary and fraught with potential harm. Gender testing was eliminated in 1992 by the International Amateur Athletic Federation, but it is still required in the Olympic Games.

Chapter References

10. Burris TP, Guo W, McCabe ER, The gene responsible for adrenal hyperplasia congenital, DAX-1, encodes a nuclear hormone receptor that defines a new class within the superfamily, Recent Prog Horm Res 51:2141, 1996.
In many societies throughout history, puberty has been a time of celebration. The changes of puberty announce the acquisition of fertility. However, it is precisely these psychological, social, and cultural forces that make this a stressful, difficult transition for many individuals. Unfortunately, the earlier age of puberty in modern times has made it difficult to cope with an earlier sexuality. Early teenage pregnancies are a relatively new problem (previously they were a biological impossibility). This change in modern life, probably due to improved nutrition and living conditions, makes it all the more important to understand puberty.

The ability to diagnose and manage disorders of female pubescence requires a thorough understanding of the physical and hormonal events that mark the evolution of the child into a sexually mature adult capable of reproduction. Abnormalities in this process of developmental endocrinology either lead to premature, attenuated, or retarded (delayed) puberty. This chapter reviews the important landmarks and mechanisms of normal female maturation as well as the abnormalities which lead to precocious or delayed puberty.

The Physiology of Puberty

The Period of Infancy and Childhood

The hypothalamus, anterior pituitary gland, and gonads of the fetus, neonate, the prepubertal infant and child are all capable of secreting hormones in adult concentrations. Even during fetal life, serum concentrations of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) reach adult levels at midgestation but fall thereafter as the high level of pregnancy steroid hormones exerts inhibitory feedback. Separation of the newborn from its sources of maternal and placental estrogen and progesterone releases newborn FSH and LH from this negative feedback. A prompt rise in gonadotropin secretion follows (along with an increase in inhibin B levels) that, in female neonates, may occasionally reach levels greater than those in the normal adult menstrual cycle. As a result, transient estradiol secretory equivalent to the level of the midfollicular phase of the menstrual cycle is induced and is associated with waves of ovarian follicle maturation and atresia. Full negative feedback is rapidly attained; ovarian steroids and gonadotropins decline and remain at very low levels until 6–8 years of age. During this period, the hypothalamic-pituitary system controlling gonadotropins (the “gonadostat”) is highly sensitive to negative feedback of estrogen (estradiol concentration in these years remains low at 10 pg/mL). Studies on gonadal dysgenesis and other hypogonadal infants indicate that the “gonadostat” is 6–15 times more sensitive to negative feedback at this period than in the adult. Therefore, gonadotropin secretion is in part restrained by even extraordinarily low levels of estrogen.

This reduction in the infant's gonadotropins is not entirely due to exquisite sensitivity to negative feedback. Low levels of FSH and LH even exist in hypogonadal children (with gonadal dysgenesis) between the ages of 5 and 11 years and are similar to the low levels in normal infants of this age. Because gonadotropin-releasing hormone (GnRH) infusion stimulates moderate LH and FSH secretion in these gonadal subjects, a central nonsteroidal suppressor of endogenous GnRH and gonadotropin synthesis appears to be operative.

Although peptide hormone concentrations are low throughout infancy and the prepubertal child, FSH and LH display evidence of pulsatile secretion. As noted, pituitary gonadotropins react with small but significant responses to exogenously administered GnRH, which although quantitatively less than those achieved in puberty are nevertheless capable of inducing the immature gonad to respond with modest steroid secretion. Immaturity of the various endocrine components is not the rate-limiting factor in the onset of puberty.

The Prepubertal Period

As puberty approaches, three critical changes in the low endocrine homeostatic function of childhood emerge:

1. **Adrenarche**.
   2. Decreasing repression of the “gonadostat.”
   3. Gradual amplification of the peptide-peptide and peptide-steroid interactions leading to “gonadarche.”

### Adrenarche

The growth of pubic and axillary hair is due to an increased production of adrenal androgens at puberty. Thus, this phase of puberty is often referred to as adrenarche (or pubarche). **Premature adrenarche** by itself is occasionally seen; i.e., pubic and axillary hair without any other sign of sexual development. **Premature thelarche** (breast development) without other signs of puberty is very rare, but does occur. Increased adrenal cortical function, expressed by a rise in circulating
dehydroepiandrosterone (DHA), dehydroepiandrosterone sulfate (DHAS), and androstenedione associated with increased adrenal 17α-hydroxylase and 17,20-lyase activity (the P450c17 enzyme), occurs progressively in late childhood from about age 6–7 to adolescence (13–15 years of age). This steroid secretion is associated with an increase in size and differentiation of the inner zone ( zona reticularis) of the cortex. Generally, the beginning of adrenarche precedes by 2 years the linear growth spurt, the rise in estrogens and gonadotropins of early puberty, and menarche at midpuberty. Because of this temporal relationship, activation of adrenal androgen secretion has been suggested as a possible initiating event in the ontogeny of the pubertal transition.

Considerable evidence, however, supports a dissociation of the control mechanisms that initiate adrenarche and those governing GnRH-pituitary-ovarian maturation ("gonadarche"). Premature adrenarche (precocious appearance of pubic and axillary hair before age 8 years) is not associated with a parallel abnormal advancement of gonadarche. In hypogonadotrop hypogonadism (gonadal dysgenesis) or in hypogonadotropic states such as Kallmann's syndrome, adrenarche occurs despite the absence of gonadarche. When adrenarche is absent, as in children with cortisol-treated Addison's disease (hypoadrenalism), gonadarche still occurs. Finally, in true precocious puberty occurring before 6 years of age, gonadarche precedes adrenarche.

Plasma levels of adrenal androgens change without corresponding changes in cortisol and ACTH during fetal life, puberty, and aging. Furthermore, in other circumstances such as chronic disease, surgical stress, recovery from secondary adrenal insufficiency, and anorexia nervosa, changes in ACTH-induced cortisol secretion are not accompanied by corresponding changes in plasma adrenal androgen levels. Thus, adrenarche does not appear to be under direct control of gonadotropins or ACTH.

A pituitary adrenal androgen stimulating factor formed by cleavage of a high molecular weight precursor, proopiomelanocortin (POMC), which also contains ACTH and b-lipotropin, acts on an ACTH prepared and maintained adrenal, has been suggested as the agent stimulating adrenarche. A large glycoprotein has been identified that also displayed adrenal androgen stimulating activity. However, in a study confirming the dissociation between plasma adrenal androgens and cortisol in children and adolescents with Cushing's disease and ectopic ACTH producing tumors, all known proopiomelanocortin-related peptides, including ACTH, b-endorphin, and b-lipotropin did not have a determinative role in the initiation of adrenarche. Studies fail to demonstrate a relationship between melatonin secretion and adrenarche.

A study of the kinetics of the 3b-hydroxysteroid dehydrogenase enzyme in human adrenal microsomes suggests that the changes in adrenal secretion from fetal life to adulthood can be explained by local steroid inhibition of key enzymes within the adrenal, acting to a variable degree in different layers of the cortex and at different stages of development. Puberty is associated with the emerging prominence of a zona reticularis (the source of DHA and DHAS) that is relatively deficient in 3b-hydroxysteroid dehydrogenase activity, a state that is maintained during adult life. The factors that regulate and suppress this enzyme activity remain unknown. Another hypothesis suggests that serine phosphorylation of P450c17 is necessary for 17,20-lyase activity, and that this phosphorylation, stimulated by a yet unknown mechanism, is responsible for the initiation of adrenarche. It is far too early to say that the factors controlling adrenarche remain obscure.

**Decreasing Repression of the "Gonadostat"**

Regardless of its relation to adrenarche, factors which induce gonadarche in late prepuberty involve derepression of the central nervous system (CNS)-pituitary gonadostat, progressive responsiveness of the anterior pituitary to exogenous (and presumably endogenous) GnRH, and follicle reactivity to FSH and LH.

For approximately 8 years, from early infancy to the prepubertal period, LH and FSH are suppressed to very low levels. The mechanisms for this restraint on gonadotropin secretion are a highly sensitive negative feedback of low level gonadal estrogen on hypothalamic and pituitary sites, and an intrinsic central inhibitory influence on GnRH that reduces basal gonadotropin concentrations even in agonal children. Gonadal dysgenesis patients display marked elevations of gonadotropins for the first 2–3 years of life. Thereafter, a striking decline in concentrations of FSH and LH occurs, reaching a nadir at 6–8 years. By age 10–11 (at the time puberty would have occurred), however, gonadotropins are elevated once again to the postmenopausal range. The overall pattern of basal gonadotropin secretion in agonal children is qualitatively similar to that observed in normal females.

Whereas negative feedback inhibition may play the more important role in early childhood, the central intrinistic inhibitor becomes functionally dominant in midchildhood and persists up to prepuberty. Suppression of, or damage to, the neural source of this inhibition has been postulated in the pathogenesis of the precocious puberty secondary to hypothalamic lesions that compress or destroy posterior hypothalamic areas. Thus, normal pubertal timing of gonadarche, with the reactivation of gonadotropin synthesis and secretion, results from the combined reduction in intrinsic suppression of GnRH and decreased sensitivity to the negative feedback of estrogen.

It has been suggested that the reversal of central intrinsic suppression is due to a reduction in melatonin secretion by the pineal gland. In lower animals affected by photoperiodicity, pineal melatonin appears to inhibit hypothalamic-pituitary gland secretion. While melatonin may play a role in the altered timing of puberty associated with pineal tumors and in the pathophysiology of central precocious puberty, there is no evidence that it is important in the physiologic onset of normal puberty in humans. In two large studies of circadian rhythms of serum melatonin from infancy to adulthood (1–18 years) the decline in the nocturnal surge of melatonin, thought to have been exclusively related to the pubertal conversion, was observed to begin in infancy and progressively decline through pubescence. Pinealectomy in agonal primates does not prevent the initiation of FSH and LH seen during transition from infancy to childhood nor the return of gonadotropins with the advent of puberty.

The fascinating search for the factor(s) involved in the derepression of the "gonadostat" so crucial to the timing of puberty continues. POMC-related peptides do not appear to change during the transitional period. The ontogeny of the GnRH gene and its expression, so elegantly demonstrated in rodents, is yet to be extended to the primate.

**Alteration and Amplification of GnRH-Gonadotropin and Gonadotropin-Ovarian Steroid Interactions**

FSH and LH levels increase during the progress through the stages of puberty. Rhythmic pulses of GnRH given to immature rhesus monkeys will initiate activity of the pituitary-gonadal apparatus, supporting the primacy of endogenous GnRH in the establishment and maintenance of puberty. Similar effects have been demonstrated in prepubertal girls. Normal pubertal maturation in girls is also accompanied by changes in the pattern of gonadotropin responses to the hypothalamic-releasing hormone GnRH. FSH responses to GnRH are initially pronounced but decrease steadily throughout the onset of puberty. In contrast, LH responses are low in prepubertal girls and increase strikingly during puberty. This is the basis of the observation that in general FSH rises initially and plateaus in midpuberty while LH tends to rise more slowly and reaches adults levels in late puberty. The increased amplitude and frequency of pulsatile GnRH are believed to provoke progressively enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of FSH and LH secretion.

The amplification of peptide-steroid interactions during pubescence is not restricted to the GnRH impact on gonadotropin or steroid feedback on the pituitary and gonadotropins. As gonadotropin secretion appears, ovarian follicle steroid synthesis is stimulated and estrogen secretion rises. This is the basis of the observation that in general FSH rises initially and plateaus in midpuberty while LH tends to rise more slowly and reaches adults levels in late puberty. The increased amplitude and frequency of pulsatile GnRH are believed to provoke progressively enhanced responses of FSH and LH secretion. GnRH acts as a self-primer on the gonadotrope cells of the anterior pituitary by inducing cell surface receptors specific for GnRH and necessary for its action (up-regulation). Thus, gonadotrope cells increase their capacity to respond to GnRH first by synthesis and later by secretion of enhanced responses of gonadotropins. As gonadotropin secretion increases, ovarian follicle steroid synthesis is stimulated and estrogen secretion rises.

Elsewhere (Chapter 5 and Chapter 6), the evidence for the dichotomous effects of estrogen feedback on the anterior pituitary has been reviewed. Suffice to say, midpuberty estrogen enhances LH secretion to GnRH (positive feedback) while combining with inhibin to maintain relative inhibition (negative feedback) of FSH response.

The amplification of peptide-steroid interactions during pubescence is not restricted to the GnRH impact on gonadotropin or steroid feedback on the pituitary and gonadotropins. As the pubertal transition advances there is a disproportionate rise of biologically potent LH beyond the increase seen in immunologic LH. This marked increase in the bioactive to immunoreactive ratio is due to molecular alterations in the glycosylation pattern of LH, as reviewed in Chapter 2 under Heterogeneity.

Prior to pubert, gonadotropin levels are low but still associated with pulses. The onset of augmented GnRH pulses first occurs during sleep. There is stepwise associated release of LH in both sexes, a consequence of LH responses to endogenous GnRH. The early stages of puberty are associated with a marked nocturnal augmentation of FSH and LH pulses (both amplitude and frequency, but especially frequency); this difference between nighttime and daytime switches by late puberty with an increase in daytime and a decrease in sleep pulsatility. It should be noted that day-night differences exist before puberty, but the differences become more marked with the onset of puberty. This change is not abrupt with the onset of puberty. Very sensitive assays can detect an increase in FSH and LH (both day and night) in the months preceding the beginning of breast development.

Sleep-related LH pulses also are seen in children with idiopathic precocious puberty, in anorexia nervosa patients during intermediate stages of exacerbation and
recovery, and also in agonadal patients during the pubertal age period when their gonadotropins are returning from midchildhood reductions. \(^4\) GnRH pulses appear and are maintained independent of steroid feedback.

## Puberty

The cascade of events initiated by the release of pulsatile GnRH from prepubertal feedback and central negative inhibition results in increased levels of gonadotropins and steroids with appearance of secondary sexual characteristics and eventual adult function (menarche and, later, ovulation). Between the ages of 10 and 16 the endocrine sequence observed includes, first, increased pulsatile patterns of LH during sleep, followed by similar pulses of less amplitude occurring throughout the 24-hour day. Episodic peaks of estradiol result and menarche appears. By mid to late puberty, maturation of the positive feedback relationship between estradiol and LH is established, leading to ovulatory cycles.

### Timing of Puberty

Although the major determinant of the timing of puberty is genetic, other factors appear to influence the time of initiation and the rate of progression of puberty: geographic location, exposure to light, general health and nutrition, and psychologic factors. For example, children with a family history of early puberty start early. \(^3\) Children closer to the equator, at lower altitudes, those in urban areas, and mildly obese children start earlier than those in Northern latitudes, at higher elevations above sea level, in rural areas, and normal weight children, respectively. There is a fairly good correlation between the times of menarche of mothers and daughters and between sisters. \(^2\) There is a correlation between age of onset and duration of puberty; the earlier the onset, the longer the duration. \(^5\)

The decline in the age of menarche displayed by children in developed countries undoubtedly reflects improved nutritional status and healthier living conditions. Frisch argued that a critical body weight (47.8 kg) must be reached by a girl to achieve menarche. \(^2\) Possibly more important than total weight is the shift in body composition to a greater percent fat (from 16.0 to 23.5%), which in turn is influenced by the nutritional state. \(^2\) Indeed, moderately obese girls (20–30% over normal weight) have earlier menarche than normal weight girls. \(^2\) Conversely, anorectics and intense exercisers (low weight or low percent fat component of weight) have delayed menarche or secondary amenorrhea. That other factors are involved is indicated by the delayed menarche experienced by morbidly obese girls (greater than 30% overweight), diabetics, and intense exercisers of normal weight. Intriguingly, blind girls experience earlier menarche. \(^2\) Furthermore, girls with idiopathic central precocious puberty may undergo menarche at a total body fat of 19%, and girls with no signs of puberty may have measured total body fat of 27%. \(^5\)

It is reasonable to hypothesize that central mechanisms bring about maturation of the hypothalamic-pituitary-ovarian axis, which in turn stimulates growth to the critical weight as well as the increases in body fat composition. However, not all auxologic studies have found a relationship between the onset of puberty and either body fat mass or body fat distribution. \(^2\) Nevertheless, the identification of leptin (discussed in Chapter 19) has revitalized the importance of a relationship between body fat and reproductive function.

Leptin is a 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 energy behavior and endocrine physiology. Several observations support a role for leptin in reproductive physiology.

1. Leptin administration accelerates the onset of puberty in rodents. \(^2\)
2. Leptin levels increase at puberty in boys. \(^4\)
3. Low leptin levels are present in athletes and in patients with anorexia, or delayed puberty. \(^4\)
4. Mice lacking leptin undergo normal sexual development, but remain prepubertal and never ovulate, but fertility is restored with leptin administration. \(^2\)

Leptin levels increase during childhood until the onset of puberty, suggesting that a threshold level of leptin (and, therefore, a critical amount of adipose tissue, the source of leptin) is necessary for puberty to begin. \(^2\) The higher the level of leptin, the earlier the age of menarche. \(^2\) Girls with idiopathic central precocious puberty have higher levels of leptin. \(^2\) These relationships support the existence of communication between the central nervous system and body fat in the process of puberty, with leptin serving as the messenger.

After puberty, leptin levels in boys return to the prepubertal range (apparent suppression by testosterone). In girls, leptin levels are higher than in boys and decrease with increasing Tanner stages of puberty. \(^2\) Thus with puberty, there is increasing sensitivity to leptin. Or, in another way to look at this relationship, a decrease in leptin during puberty may allow greater food intake for growth by lowering the satiety signal.

## Stages of Pubertal Development

On the average, the pubertal sequence of accelerated growth, development, adrenarche, and menarche requires a period of 4.5 years (range 1.5 to 6 years). The largest body of data was accumulated in healthy European girls; current North American standards are approximately 6 months earlier for each stage. \(^2\) Secondary sex characteristics develop earlier in black girls than in white girls.

In the early 1990s, the American Academy of Pediatrics performed a cross-sectional study of pubertal development in American girls seen in pediatric office practices. Although this study was handicapped by deriving data from office practices rather than a randomized population sampling, it is the largest study available, containing 17,077 girls (9.6% black Americans and 90.4% white). The important findings are as follows:

<table>
<thead>
<tr>
<th>Stages of Pubertal Development</th>
<th>Black Americans</th>
<th>White Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast and/or public hair at age 7:</td>
<td>27.2%</td>
<td>6.7%</td>
</tr>
<tr>
<td>at age 8:</td>
<td>48.3%</td>
<td>14.7%</td>
</tr>
<tr>
<td>Menarche at age 11:</td>
<td>27.9%</td>
<td>13.4%</td>
</tr>
<tr>
<td>at age 12:</td>
<td>62.1%</td>
<td>35.2%</td>
</tr>
<tr>
<td>Mean age for thelarche:</td>
<td>8.87 years</td>
<td>9.96 years</td>
</tr>
<tr>
<td>Mean age for adrenarche:</td>
<td>8.78 years</td>
<td>10.51 years</td>
</tr>
<tr>
<td>Mean age for menarche:</td>
<td>12.16 years</td>
<td>12.88 years</td>
</tr>
</tbody>
</table>

The data from the American Academy of Pediatrics study indicates that a substantial number of girls display development 6 months to one year earlier than previously documented, although the age of menarche has remained stable. On average, black American girls begin puberty between ages 8 and 9, and white American girls by age 10.

In general, the first sign of puberty is an acceleration of growth followed by breast budding (thelarche). Breast development follows a well-recognized sequence of events. Breast budding is a change distinguished by enlargement and elevation of the nipples and areolae. This is followed by elevation of the breast by the building of the breast mound. Just prior to the formation of the final adult contours, the areolae form secondary mounds.

Although the sequence may be reversed, adrenarche usually appears after the breast bud with axillary hair growth 2 years later. In approximately 20% of children, pubic hair growth is the first sign of puberty. Menarche is a late event, occurring after the peak of growth has passed.
Growth

An adolescent girl’s growth spurt occurs 2 years earlier (at 11–12 years) than that of a boy, and in 1 year, her rate of growth doubles, yielding a height increment of between 6 and 11 cm (2.4–4.3 inches). The average girl reaches this growth peak about 2 years after breast budding and 1 year prior to menarche. Evidence suggests that growth acceleration is due to estrogen and concomitant increases in growth hormone production and secondary stimulation of insulin-like growth factor-I (IGF-I) levels. Adrenal androgens are not involved because cortisol-repleted Addisonian patients display normal pubertal growth patterns.

In a remarkable study of African pygmies, it was discovered that the short stature of adult pygmies is due primarily to a failure of growth to accelerate during puberty, and that the principal factor responsible for normal pubertal growth is insulin-like growth factor-I (IGF-I). Growth hormone exerts its action through a locally produced mediator, insulin-like growth factor-I. In addition, growth hormone can directly stimulate epiphyseal cartilage growth. Normal growth at puberty requires the concerted action of growth hormone, insulin-like growth factor-I, and sex steroids. The increase in circulating insulin-like growth factor-I at puberty correlates with sexual development and results from the interaction between sex steroids and growth hormone. Specifically, the increase in sex steroids in turn increases the secretion of growth hormone, which stimulates the production of insulin-like growth factor-I. However, studies also indicate that the sex steroids can have a direct effect on bone growth independent of growth hormone. Thus, Laron-type dwarfs (who have a genetic defect in the growth hormone receptor and cannot stimulate IGF-I secretion) can undergo a growth spurt at puberty in response to the sex steroids. However, normal pubertal growth velocities require the combined action of the sex steroids and growth hormone. The sex steroid hormones also limit the ultimate height attained by stimulating epiphyseal fusion. The principal sex steroid involved in pubertal growth is probably estrogen in both boys and girls, with aromatization of androgens being the source of estrogen in boys.

The most abundant hormone produced by the pituitary gland is growth hormone, which is secreted not as a single substance but as one predominant form and one smaller variant. Growth hormone is encoded by 5 genes located on chromosome 17q22-q24. One gene is for the predominant form in the pituitary; 3 of the genes are expressed in the placenta. The pituitary gene is regulated by growth hormone-releasing hormone, thyroid hormone, and glucocorticoids. Besides the stimulation of IGF-I in cartilage, growth hormone also stimulates IGF-I production in a variety of tissues throughout the body, especially in the liver (the main source of circulating IGF-I).

The hepatic production of insulin-like growth factor binding protein-1 (IGFBP-1) is regulated by insulin. The circulating levels of IGFBP-1 decrease throughout puberty, in response to the relative hyperinsulinemia that accompanies puberty in response to increasing insulin resistance. This change would allow greater metabolic activity of IGF-I, an important mediator of growth.

Like the gonadotropins, growth hormone is secreted in pulsatile fashion, and during puberty, the amplitude of the pulses increases, especially during sleep. Your grandmother was right when she said: sleep and you’ll grow. The age at which an increase in pulse amplitude first occurs corresponds to the age of most rapid growth. The growth response to growth hormone is correlated with the increases in pulse amplitude, not a change in baseline levels. Slower growing children display fewer and smaller pulses of growth hormone. The pulsatile pattern of growth hormone secretion is regulated by stimulation from growth hormone-releasing hormone and inhibition from somatotropin release-inhibiting hormone, both released into the hypothalamic-pituitary portal circulation from hypothalamic nuclei. This mechanism is influenced at multiple levels by estrogens and androgens. Prior to puberty, the sex steroid hormones are not involved with growth hormone secretion, beyond a low maintenance effect on secretion. At puberty, however, the dynamics of growth hormone secretion are critically dependent on the gonadal sex steroid hormones. Growth hormone secretion must be very sensitive to the stimulatory effect of estrogens because growth hormone levels increase before any signs of sexual development appear.

The amounts of estrogen required to stimulate long-bone cortical growth are incredibly small. Doses of 100 nanograms of estradiol per kilogram body weight per day increase the amplitude of growth hormone pulsatile secretion and produce maximal growth in gonadectomized recipients. These doses are insufficient to cause breast budding, vaginal cornification, or an increase in sex hormone-binding globulin. These low dose effects are consistent with the observation that girls attain peak height velocity early in puberty at a serum estradiol concentration of 20 pg/mL, which is one-sixth the mean level of adult women. Furthermore, at low doses, estrogen stimulates growth hormone-induced IGF-I secretion, while high doses suppress IGF-I levels.

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. 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. And most impressively, a man with an aromatase deficiency, treated with estrogen, demonstrated that both androgens and estrogens are necessary in order to for males to reach optimal bone mass.

Osteoporosis and vertebral fractures are less common in black than in white women. Vertebral bone density increases rapidly and significantly during adolescence, and the increase is greater in black girls, providing one explanation for the racial difference in osteoporosis. The pubertal increase in bone density ranges from 10% to 20%, an accumulation which provides 10–20 years of protection against the normal age-related loss of skeletal mass. Calcium supplementation during adolescence results in a significant increase in bone density and skeletal mass, providing even greater protection against future osteoporosis. Optimal growth has both immediate and long-term consequences. Adolescents with abnormal menstrual function (suppressed estrogen levels) should not be ignored, but properly evaluated and treated. The influence of the sex steroid levels on bone mass is underscored by the fact that almost all of the bone mass in the hip and the vertebral bodies will be accumulated in young women by late adolescence (age 18), and the years immediately following menarche are especially important.

Menarche

As mentioned previously, environmental factors are important in the onset of puberty. Improved living standards and nutrition in the mother antenatally, and in children postnatally, have played a significant role in producing taller, heavier children with earlier maturation. Studies of identical twins and nonidentical twins indicate that the age at menarche is chiefly controlled by genetic factors when the environment is optimal. In affluent cultures, the trend toward lowering of the menarcheal age and puberty halted around 1960. In the 1700s, the mean age of voice change in the Boys’ Bach Choir in Leipzig was 18, now it is 13.5 years. Recent studies have indicated a very slight upward trend in the age of menarche, perhaps a response to some environmental deterioration.
The relationship between menarche and the growth spurt is relatively fixed: menarche occurs after the peak in growth velocity has passed. Hence slower growth, totaling no more than 6 cm (2.4 inches) is noted after initiation of menses.

The normal age range of menarche in U.S. girls is 9.1–17.7 years with a median of 12.8. The final endocrine hallmark of puberty is the development of positive estrogen feedback on the pituitary and hypothalamus. This feedback stimulates the midcycle surge of LH required for ovulation. Thus, the menses following menarche are usually anovular, irregular, and occasionally heavy. Anovulation lasts as long as 12–18 months after menarche, but there are reports of pregnancy before menarche. Ovulation increases in frequency as puberty progresses, but it is common for 25–50% of adolescents to still be anovulatory 4 years after menarche.

Summary of Pubertal Events

The onset of puberty is an evolving sequence of maturational steps. The hypothalamic-pituitary-gonadal system differentiates and functions during fetal life and early infancy. Thereafter, it is suppressed to low activity levels during childhood by a combination of hypersensitivity of the “gonadostat” to estrogen negative feedback and an intrinsic CNS inhibitor. All the components located below GnRH (below the CNS) are competent to respond at all ages (as will be seen in the pathogenesis of precocious puberty). After a decade of functional GnRH insufficiency (between late infancy and the onset of puberty), GnRH secretion is resumed and gonadarche (the reactivation of the CNS-pituitary-ovarian apparatus) appears. Prolongation of intrinsic CNS suppression or disability in any of the components of the gonadarche cascade leads to delayed or absent pubescence.

1. FSH and then LH levels rise moderately before the age of 10 and are followed by a rise in estradiol. An increase in LH pulses is first seen only in sleep but gradually extends throughout the day. In the adult, they occur at roughly 1.5–2 hourly intervals.

2. As gonadal estrogen increases (gonadarche), breast development, female fat distribution, and vaginal and uterine growth occur. Skeletal growth rapidly increases as a result of initial gonadal secretion of low levels of estrogen, which increases the secretion of growth hormone, which in turn stimulates the production of IGF-I.

3. Adrenal androgen (adrenarche) and, to a lesser degree, gonadal androgen secretion cause pubic and axillary hair growth. Adrenarche plays little, if any, part in skeletal growth. While temporarily related to gonadarche, adrenarche is an independent, functionally unrelated biological event.

4. At midpuberty, sufficient gonadal estrogen secretion proliferates the endometrium, and the first menses (menarche) occurs.

5. Postmenarcheal cycles are initially anovulatory. Sustained, predictable positive LH surge responses to estradiol with ovulation are late pubertal events.

Blood Hormone Concentrations During Female Puberty

<table>
<thead>
<tr>
<th>Tanner Stage</th>
<th>FSH IU/L</th>
<th>LH IU/L</th>
<th>Estradiol pg/mL</th>
<th>DHA ng/dL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>0.9–5.1</td>
<td>1.8–9.2</td>
<td>&lt;10</td>
<td>19–302</td>
</tr>
<tr>
<td>Stage 2</td>
<td>1.4–7.0</td>
<td>2.0–16.6</td>
<td>7–37</td>
<td>45–1904</td>
</tr>
<tr>
<td>Stage 3</td>
<td>2.4–7.7</td>
<td>5.6–13.6</td>
<td>9–59</td>
<td>125–1730</td>
</tr>
<tr>
<td>Stage 4</td>
<td>1.5–11.2</td>
<td>7–14.4</td>
<td>10–156</td>
<td>153–1321</td>
</tr>
<tr>
<td>Adult follicular</td>
<td>3–20</td>
<td>5–25</td>
<td>30–100</td>
<td>162–1620</td>
</tr>
</tbody>
</table>

Precocious Puberty

Puberty is the biologic transition between immature and adult reproductive function. Its timing, endocrine milieu, and physical expressions have been characterized sufficiently to set clinically reasonable time limits for the normal appearance of female maturity and to allow recognition of the pathogenesis and pathophysiology of most of the causes of premature or delayed pubescence.

If one accepts the mean ±2.5 standard deviations as encompassing the normal range, then pubertal changes before the age of 8 are regarded as precocious. However, the most recent cross-sectional study (by the American Academy of Pediatrics) indicates that a substantial number of normal girls begin pubertal development before age 8.5. Certainly, thelarche and adrenarche before age 6 deserves evaluation. From age 6 to age 8, the clinician’s response will be influenced by the anxiety and distress in patient and parents, and the clinical presentation. Increased growth is often the first change in precocious puberty. This is usually followed by breast development and growth of pubic hair. On occasion, adrenarche, thelarche, and linear growth occur simultaneously. Menarche, however, can be the first sign.

Traditionally, precocious puberty has been divided into two classifications:

1. GnRH-Dependent Precocious Puberty. Complete, isosexual, central (or specifically GnRH- and gonadotropin-dependent) precocity — also known as true precocious puberty. These terms all refer to early activation of the hypothalamic-pituitary-gonadal axis.
2. GnRH-Independent Precocious Puberty. Incomplete, isosexual or heterosexual, peripheral or precocious pseudopuberty. Sexual maturation in these instances may be due to extra pituitary secretion of human chorionic gonadotropin (HCG) or sex steroid secretion independent of hypothalamic-pituitary gonadotropin stimulation. Thus, this mechanism is GnRH-independent.

In clinical practice, these classifications are of little practical use. Precocity occurs in girls 5 times more frequently than boys, and almost three-quarters of precocity in girls is idiopathic. Nevertheless, in the face of any precocious development, the clinician is obligated to rule out a serious disease process in central or peripheral sites. In girls over 4 years old a specific etiology is rarely found. In younger girls, a CNS lesion is usually present.

A classification of sexual precocity is presented to provide a guide to the possible conditions and the relative incidences encountered. It should be noted that a GnRH-dependent cause is found in up to 80% of girls, with a lesser occurrence in boys. Ovarian tumors, adrenal disease, and the McCune-Albright syndrome make up the majority of noncental precocity in girls.

### Classification and Relative Occurrence of Precocious Puberty

<table>
<thead>
<tr>
<th>Classification</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnRH-Dependent (True Precocity)</td>
<td>74.0%</td>
<td>41.0%</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>7.0%</td>
<td>26.0%</td>
</tr>
<tr>
<td>CNS problem</td>
<td>11.0%</td>
<td>—</td>
</tr>
<tr>
<td>Testicular</td>
<td>5.0%</td>
<td>1.0%</td>
</tr>
<tr>
<td>McCune-Albright syndrome</td>
<td>1.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Adrenal feminizing</td>
<td>1.0%</td>
<td>22.0%</td>
</tr>
<tr>
<td>Ectopic gonadotropin production</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

GnRH-Dependent Precocious Puberty (Precocious Development Due to Stimulation of Gonadotropin Secretion, True Precocity)

The signs of constitutional sexual precocity are due to premature maturation of the hypothalamic-pituitary-ovarian axis, resulting in production of gonadotropins and sex steroids. These patients experience an increase in growth that is associated with pubertal levels of insulin-like growth factor-I. Constitutional precocity runs in families and usually occurs very close to the “borderline” age of 8 years. On the other hand, idiopathic precocious puberty does not run in families and occurs much earlier in childhood. It must be reemphasized that these benign diagnoses should be made only by exclusion and deserve long-term follow-up, because cerebral abnormalities may not become apparent until adulthood.

Clinical presentation of true precocity may not follow the usual progression of breast and pubic hair growth, acceleration of growth rate and then menses. It is not unusual for adrenarche or menarche to be the first sign (or an adult body odor) with others following. This progression is variable, usually slower in idiopathic cases, but telescoped in precocity due to central disease.

Sexual precocity is consistent with normal reproductive life and it is not associated with premature menopause. The most serious effect of precocity is the resultant adult short stature. Because the skeleton is more sensitive to even the lowest levels of estrogen, these children are transiently tall for their age, but as a result of early epiphyseal fusion, eventually short stature results. Fifty percent are less than 5 feet tall (152 cm).

Intelectual and psychosocial development are also commensurate with chronologic age rather than stage of puberty. Expectations of emotional, social, sexual, and intellectual competence corresponding to their pubertal state leave these youngsters and their families with potentially serious difficulties on all levels of social and emotional function.

A number of CNS problems, including abnormal skull development due to rickets, can cause true precocious development. Various tumors can induce precocity, including hamartomas in the hypothalamus (the most common lesion in very young girls), craniopharyngioma, astrocytoma, glioma, neurofibroma, ependymoma, and suprasellar teratoma — all usually near the hypothalamus. Pineal tumors, for unknown reasons, have been seen only in male precocious puberty. Nontumorous causes include encephalitis, meningitis, hydrocephalus, and von Recklinghausen's disease. An injury to the skull may stimulate sexual development. The mechanism is unknown, and a latent period of 1–2 months is usually seen. A hamartoma is a hyperplastic congenital malformation in the floor of the third ventricle that usually produces precocity in the first few years of life; magnetic resonance imaging is the most sensitive method for the detection of small tumors like a hamartoma. Patients with true precocious puberty and known CNS lesions or a history of cranial irradiation should be evaluated for growth hormone deficiency because of the recognized association of these defects.

There is no unifying pathophysiologic mechanism linking this diverse spectrum of etiologies for central precocity. Increased intracerebral pressure and a predilection for intrauterine activation of gene expression in glial cells presents an intriguing model in that TGF-a stimulates GnRH release.

Ectopic gonadotropin production is a rare cause of sexual precocity accounting for less than 0.5% of cases. The most common tumors producing human chorionic gonadotropin (HCG) are chorioepitheloma and dysgerminoma of the ovary, and liver hepatoblastoma. Tumor spread may be present at the time of pubertal development, pelvic and abdominal masses accompanied by ascites are usually detectable.

True sexual precocity occurs in a small number of children with long-standing hypothyroidism. There is evidence to support the possibility that high levels of thyroid-stimulating hormone (TSH) can stimulate the FSH receptor. In addition to short stature (but not bone age acceleration), galactorrhea may be present. The sella turcica is frequently enlarged, but with thyroid replacement pubertal development will stop and even regress. The sella films will return to normal. Although reported cases have been severe and, therefore, clinically obvious, laboratory evaluation of thyroid function is indicated in all cases of sexual precocity.

### GnRH-Independent Precocious Puberty (Development Due to Availability of Sex Steroids, Precocious Pseudopuberty)

Eleven percent of girls with precocious puberty have an ovarian tumor. The tumor is usually an estrogen-producing neoplasm or cyst. Five percent of granulosa cell tumors and 1% of theca cell tumors occur before puberty. However, gonadoblastomas, teratomas, lipid cell tumors, cystadenomas and even ovarian cancers have been reported as causes of precocity. Bleeding is irregular and menorrhagic — clearly anovulatory. A pelvic mass is readily palpable in 80% of cases. The palpation of a pelvic or abdominal mass demands surgical exploration. Increasing use has been made of pelvic ultrasonography and whole body (abdominal) imaging for the
work-up of precocious puberty. In addition to estrogen and androgens, these tumors can secrete HCG.

A feminizing adrenal tumor is very rare (1% of cases) and is associated with increased blood levels of DHAS.

Drug ingestion should be suspected in all cases of precocity, especially when there is dark pigmentation of the nipples and breast areolae, an effect of certain synthetic estrogens such as stilbestrol. Common sources are oral contraceptives, anabolic steroids, and hair or facial creams.

McCune-Albright syndrome (polyostotic fibrous dysplasia) accounts for 5% of female precocity and consists of multiple disseminated cystic bone lesions that easily fracture, cafe-au-lait skin spots of various sizes and shapes, and sexual precocity. In addition, this syndrome can be associated with ovarian cysts, growth hormone- and prolactin-secreting adenomas, hyperthyroidism, adrenal hypercortisolism, and osteomalacia. Premature menarche may be the first sign of the syndrome. Skeletal abnormalities may become evident following the onset of puberty. The combination of multiple bone fractures, cafe au lait patches, and premature development should lead to the diagnosis. But remember that the manifestations of this syndrome can be varied and sometimes subtle. A technetium-99 bone scan may be necessary to demonstrate the areas of bony fibrous dysplasia.

Sexual precocity in McCune-Albright syndrome is now demonstrated to be the result of autonomous early production of estrogen by the ovaries. FSH and LH levels are low, respond poorly to GnRH stimulation, and there is an absence of nocturnal gonadotropin pulsations (all unlike central precocity). In addition, Cushings disease, acromegaly, hyperparathyroidism, and hyperthyroidism have been reported in this syndrome. The protean manifestations of this disorder suggested that the pathophysiology results from a basic defect in cellular regulation at the level of the G protein-cAMP-kinase function in affected tissues. A mutation in the alpha-subunit (Gαs) of the G protein that stimulates cAMP formation (as described in Chapter 2) has been identified in all affected tissues in patients with McCune-Albright syndrome. This mutation attenuates GTPase activity that is necessary to terminate adenylate cyclase activation; thus, affected tissues have autonomous activity. Somatic mosaicism of the alpha-subunit accounts for the fact that this mutation is not lethal and for the variation in site and activity throughout the body. This mutation may also occur in nonendocrine tissues in patients with McCune-Albright syndrome, thus explaining the occurrence of hepatitis, intestinal polyps, and cardiac arrhythmias. It is possible that this mechanism is responsible for childhood diseases other than McCune-Albright syndrome. For this reason, it has been suggested that this genetic disorder should be called inherited Gαs deficiency. Eventual fertility is unimpaired, and adult height is usually normal. These positive factors must be considered in the choice of management of the syndrome. In keeping with the autonomous nature of the gonadal activity, treatment with a GnRH agonist fails to suppress gonadal hormone secretion or reverse the sexual precocity. 97

Familial male precocious puberty is inherited in an autosomal dominant fashion and is caused by mutations in the LH receptor gene that lead to activation. Thus far, no effects in females of activating mutations of either the LH receptor gene or the FSH receptor gene have been reported.

Yet another example of precocity due to GnRH-independent gonadal secretion of estrogen is by autonomous benign ovarian follicular or luteal cysts. These children demonstrate an absence of gonadotropin pulsations, variable responses to GnRH, and a lack of suppression of puberty by a long-acting GnRH agonist. The cysts may regress and involute and then recur so that signs of sexual precocity and vaginal bleeding remit and exacerbate. The cysts are usually large and therefore palpable. GnRH testing is useful in differentiating the autonomous (nonreactive) cyst from those secondary to the FSH and LH stimulation of central true precocity (reactive).

It is now understood that nearly every cause of peripheral precocious puberty may secondarily activate the hypothalamic-pituitary-gonadal axis with development and superimposition of a central GnRH-dependent true precocity process. Presumably the central mechanism controlling the onset of puberty can be activated once a critical threshold of somatic development has been achieved by the premature production of estrogen regardless of source of secretion. This explains the previously paradoxical observations of continuing progressive puberty despite effective treatment of specific causative disease as well as the variable effectiveness of GnRH agonist therapy in McCune-Albright and the syndrome of recurrent ovarian cyst formation. GnRH testing to determine the level of activation will dictate the need for additional GnRH agonist suppression, inhibition of steroidogenesis, or peripheral aromatase inhibition therapy in these cases.

<table>
<thead>
<tr>
<th>Laboratory Findings in Disorders Producing Precocious Puberty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gonadal Size</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Idiopathic Increased</td>
</tr>
<tr>
<td>Cerebral Increased</td>
</tr>
<tr>
<td>Gonad Unilat. incr.</td>
</tr>
<tr>
<td>Albright Increased</td>
</tr>
<tr>
<td>Adrenal Small</td>
</tr>
</tbody>
</table>

### Special Cases of Precocious Development

Special cases of precocious development include the isolated appearance of one sexual characteristic: premature adrenarche or pubarche (pubic hair), premature thelarche (breast development), or premature isolated menarche. These cases present a special dilemma for the concerned clinician. Are they benign, self-limited variants of normal sexual development that do not require treatment or the first sign of a potentially accelerating process in which early exhaustive diagnosis and long-term therapy may be necessary?

Typically linear growth and skeletal maturation are not advanced, and baseline hormone levels are normal for age and sexual development. However, there is a spectrum of conditions of premature sexual maturation defined by idiopathic central precocious puberty at one end and isolated premature thelarche or premature menarche at the other. Between these poles are examples of mixed atypical problems with variable tendencies for stability or progression. Guidelines are emerging that address the issues of when sophisticated diagnostic testing (ACTH stimulation, GnRH testing, ultrasonography, bone assessments) is indicated.

### Premature Thelarche

Premature thelarche usually occurs in the first few years of life and is usually self-limited, requiring no therapy. Follow-up has revealed that these children experience normal puberty and growth and, eventually, normal reproduction. The breast growth may regress after a few months, wax and wane for years, or last until puberty. Premature thelarche can be unilateral.

### Premature Menarche

Isolated premature menarche without other evidence of maturation is an exceedingly rare presentation of precocity, infection, the presence of a foreign body, abuse and trauma, and local neoplasms should be considered. Linear growth, development, and fertility are not affected.

### Premature Adrenarche

Premature adrenarche, the appearance of pubic hair before age 8 in girls, is the consequence of an early modest increase in the adrenal androgens, androstenedione, dehydroepiandrosterone and dehydroepiandrosterone sulfate. Normal levels according to stage of puberty are available in the literature. An adrenal enzyme deficit should be excluded by appropriate laboratory testing, but it is rarely discovered in a prepubertal child who presents only with early growth of pubic hair. Thus, an ACTH stimulation test is not necessary; a measurement of the circulating early morning level of 17-hydroxyprogesterone will suffice. Treatment is not necessary because the transient acceleration in growth and bone maturation is no major influence on puberty or final height. Surveillance of these patients should be continued because, although not certain, it has been suggested that they have an increased incidence of anovulation, hirsutism, and hyperinsulinemia. Sparse hair growth on the vulva does not represent precocious pubarche.

An ACTH stimulation test should be performed in patients who have an advanced bone age and circulating androgen levels that are greater than those seen in the early stages of puberty. ACTH testing of all children with premature adrenarche will yield results in some that are consistent with mild errors of steroidogenesis.
However, even if present, these mild enzyme changes do not require treatment, and hence, exact diagnosis is not necessary. Treatment is indicated only for unequivocal cases of 21-hydroxylase deficiency, in whom baseline 17-hydroxyprogesterone levels will be diagnostic.

<table>
<thead>
<tr>
<th>Tanner Staging</th>
<th>Breast</th>
<th>Pubic Hair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1 (prepubertal)</td>
<td>Elevation of papilla only</td>
<td>No pubic hair</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Elevation of breast and papilla as small mound, areola diameter enlarged. Median age: 9.3 years</td>
<td>Sparse, long, pigmented hair chiefly along labia majora. Median age: 10.5 years</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Further enlargement without separation of breast and areola. Median age: 11.2 years</td>
<td>Dark, coarse, curled hair sparsely spread over mons. Median age: 11.4 years</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Secondary mound of areola and papilla above the breast. Median age: 12.1 years</td>
<td>Adult-type hair, abundant but limited to the mons. Median age: 12.0 years</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Recession of areola to contour of breast. Median age: 14.6 years</td>
<td>Adult-type spread in quantity and distribution. Median age: 13.7 years</td>
</tr>
</tbody>
</table>

**Diagnosis of Precocious Puberty**

The cause of precocious development may be obvious by findings in the history or physical examination. Familial occurrence helps to exclude certain disease processes (tumors). Clinically, the nature of precocity dictates certain diagnostic priorities.

1. **Rule out life-threatening disease.** This includes neoplasms of the CNS, ovary, and adrenal.
2. **Define the velocity of the process.** Is it progressing or stabilized? Management decisions hinge on this determination. Isolated, nonendocrine causes of vaginal bleeding (trauma, foreign body, vaginitis, genital neoplasm) must be excluded.

**Differential Diagnostic Steps**

**Physical Diagnosis:**
- Record of growth, Tanner stages, height and weight percentiles.
- External genitalia changes.
- Abdominal, pelvic, neurologic examination.
- Signs of androgenization.
- Special findings: McCune-Albright, hypothyroidism.

**Laboratory Diagnosis:**
- Bone age.
- Head CT scan or MRI, ultrasonography of abdomen and pelvis.
- FSH, LH, HCG assay.
- Thyroid function tests (TSH and free $T_4$).
- Steroids (serum DHAS, testosterone, estradiol, progesterone, 17-hydroxyprogesterone).
- GnRH testing.

If the full signs of sexual precocity are present, and basal or GnRH-stimulated gonadotropins are in the pubertal range, a pituitary source of gonadotropins is suspected. A simplified GnRH test can be used as follows: 

[111]
GnRH, 100 µg, administered subcutaneously.
Serum LH level 40 minutes after injection should be less than 8 IU/L.

Any abnormality on neurologic exam or imaging points toward central (CNS in origin) precocious puberty. If these are all normal, idiopathic sexual precocity is the most likely diagnosis. It should be emphasized that basal serum gonadotropins may be in the prepubertal range in the early stages of idiopathic or central precocious puberty; with time and progression of sexual development these will rise to the pubertal range. However, an ectopic source of HCG should be considered if serum gonadotropins are suppressed while estradiol is markedly elevated; a situation easily confirmed by an immunoassay specific for the b-subunit of HCG. The rare feminizing adrenal tumor may be present if the laboratory picture is more one of elevated adrenal androgens with only slightly elevated serum estradiol and suppressed serum gonadotropins. Abdominal and pelvic ultrasonography or magnetic resonance imaging (MRI) is indicated.

When signs of sexual precocity are associated with accelerated growth and skeletal maturation, in the absence of virilization, the etiology may be an ovarian tumor or cyst. A pelvic mass is usually palpable. In this situation, serum FSH and LH are suppressed, while serum estradiol is usually elevated. An elevated serum progesterone suggests an ovarian luteoma. Pelvic ultrasound or imaging can help to confirm the presence of an ovarian mass. Laparotomy is indicated to confirm the diagnosis and carry out surgical resection.

Adrenal hyperplasia or a virilizing adrenal or ovarian tumor must be considered if signs of sexual precocity are accompanied by virilization. With elevation of serum 17-hydroxyprogesterone (17-OHP) and adrenal androgens, the diagnosis of 21-hydroxylase deficient adrenal hyperplasia is established, whereas an elevation of serum 11-deoxycortisol leads to the diagnosis of 11b-hydroxylase deficient adrenal hyperplasia. If these two serum hormones are normal, while serum DHA5 or androstenedione is elevated, an adrenal tumor or a virilizing ovarian tumor is suspect. Ultrasound examination and abdominal imaging can be utilized to further localize the tumor.

Breast development usually correlates with a bone age of 11 and menarche with a bone age of 13. If breast and genital development, pubic hair growth, and vaginal bleeding are seen in a short child with a delayed bone age, primary hypothyroidism is the most likely diagnosis. This can be confirmed by finding a low serum T₄ and elevated TSH concentration. Serum FSH and LH levels may be in the pubertal range, but these will decrease following thyroid treatment. Galactorrhea may be present along with elevated serum prolactin concentrations. These return to normal with thyroid treatment.

### Treatment of Precocious Development

The objectives of management and treatment of precocious puberty include:

1. Diagnose and treat intracranial disease.
2. Arrest maturation until normal pubertal age.
3. Attenuate and diminish established precocious characteristics.
4. Maximize eventual adult height.
5. Avoidance of abuse, reduction of emotional problems, and contraception if necessary.

A number of therapies have been used to achieve these goals. These have included medroxyprogesterone acetate, cyproterone acetate, and danazol. In addition to undesirable side effects, bone maturation and growth were not regularly or sufficiently controlled. Major progress has been made with the use of GnRH analogues for the treatment of true precocious puberty.

The short half-life of GnRH is due to rapid cleavage of bonds between amino acids 5–6, 6–7, and 9–10. Substitution of amino acids at position 6 and replacement of the C-terminal glycine amide has produced effective GnRH agonists. Agents can be chosen that are administered subcutaneously, intranasally daily, or in long-acting depot forms.

After an initial short-term “flare” stimulation of gonadotropin release, desensitization and down-regulation follow, yielding profound reduction in gonadotropins, steroid production, and biologic effects. Substantial regression of pubertal characteristics, amenorrhea, and reduction in growth velocity are rapidly achieved and maintained within the first year of treatment. Final bone height is increased but is dependent upon the stage at which medication is begun, the bone age at which the drug is stopped, and the adequacy of the dose regimen.

Even individuals with advanced bone ages will achieve greater growth because suppression of gonadal steroids will delay epiphyseal fusion and prolong the duration of growth. Treatment is more effective if begun before bone age exceeds 12 years. With idiopathic GnRH-dependent precocious puberty, height predictions are more accurate using the Bayley-Prineau tables for average girls, despite advanced bone ages in patients. Some patients will demonstrate a marked slowing of growth with GNH agonist treatment, and in these patients, the addition of growth hormone will produce an excellent growth response.

### GnRH Agonists in Clinical Use

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native GnRH</td>
<td>pGluHisTrpSerTyrGly</td>
<td>Leu</td>
<td>Arg</td>
<td>ProGly-NH$_\text{p}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buserelin</td>
<td>D-Leu</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nafarelin</td>
<td>D-Ser (tertiary butanol)</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histrelin</td>
<td>D-His (tertiary benzyl)</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goserelin</td>
<td>D-Ser (tertiary butanol)</td>
<td>Aza-Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deslorelin</td>
<td>D-Trp</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptorelin</td>
<td>D-Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The dose of GnRH agonist treatment can be monitored by measuring estradiol levels. Because estradiol is the hormone that triggers growth and development, the objective is to maintain an estradiol less than 10 pg/mL, a prepubertal range. Because many commercial estradiol assays lack sensitivity in this range, it may be necessary to confirm adequate suppression by demonstrating a lack of gonadotropin response to the administration of GnRH. In general, children require higher doses of GnRH agonists to achieve suppression compared to adults. Even with treatment, adenarche will probably continue, true to its independent control system.

Sustained release pellets (goserelin) or sustained release injections (leuprolide) allow once a month dosing. Treatment is maintained until the epiphyses are fused or until appropriate pubertal and chronological ages are matched. Discontinuation of therapy is followed by prompt reactivation of the pubertal process and the development of regular ovulatory function in a pattern similar to that of normal adolescents. GnRH agonist treatment is also recommended for GnRH-secreting hamartomas of the hypothalamus. The progress of the tumor can be monitored by imaging, and risky surgery can be avoided.

GnRH agonist treatment is not effective for nonconventional forms of precocious puberty such as McCune-Albright syndrome, GnRH-independent sexual precocity, or congenital adrenal hyperplasia. However, should patients with McCune-Albright syndrome or congenital adrenal hyperplasia mature their hypothalamic-pituitary-gonadal axis and develop true sexual precocity, then supplementary GnRH agonist therapy is helpful. Primary treatment in these cases is directed toward suppression of gonadal steroidogenesis. Medroxyprogesterone acetate can be utilized in depot form to suppress LH secretion, or testolactone, an aromatase inhibitor, can be administered.

If a specific etiology for precocious puberty is identified, treatment is aimed at curing the underlying disorder. Neurosurgical excision of hypothalamic, pituitary, cerebral, or pineal tumors must be individualized in each patient. If these tumors are small and do not extend around or into vital brain structures, their removal may be successful. If complete surgical excision is not possible, radiation therapy should be considered. Although many tumors are said not to be radiosensitive, this may
Delayed Puberty

Since there is such a wide variation in normal development it is difficult to define the patient with abnormally delayed sexual maturation. Nearly all U.S. white girls and all U.S. black girls have entered puberty by age 13. However, some evaluation is needed whenever a patient and parents are concerned enough to seek a clinician's advice. Patients who have not developed signs of puberty by age 17 are very likely to have a specific problem and not physiological delay of puberty.

Delayed puberty is a rare condition in girls, and a genetic problem or hypothalamic-pituitary disorder must be suspected. In addition, anatomic abnormalities of the target organ (uterus and endometrium) or outflow tract are unique but important elements to consider in amenorrheic but otherwise normal pubertal adolescents.

The history and physical examination are very useful in the diagnostic work-up of delayed puberty. Special note should be taken of past general health, height and weight records, and the height and pubertal milestone experience of older siblings and parents, and relevant behavior such as extreme exercise or abnormal eating habits. Physiological delayed puberty tends to be familial. On physical examination, in addition to body measurements and Tanner staging of any secondary sexual characteristics present, a search for signs of hypothyroidism, gonadal dysgenesis, hypopituitarism, or chronic illness should be made. Persistent deciduous teeth are typical of hypothyroidism. The absence of pubic hair in a patient with a uterus and vagina indicates hypopituitarism. The absence of pubic hair in a patient with a vaginal pouch indicates that the patient has androgen insensitivity syndrome.

The failure of growth in stature suggests several possibilities. Isolated growth hormone deficiency is associated with somewhat delayed sexual maturation. Menarche may eventually occur, albeit delayed, but with bone age still several years below chronologic age. More global pituitary hormone deficiency will result in total pubertal delay. Finally, gonadal dysgenesis (45,X) will be associated with decreased height and sexual infantilism with normal to slightly reduced bone age and hypergonadotropism.

Neurologic examination is important; evidence of intracranial disease, restricted visual fields, or absent sense of smell are key findings. Anatomic defects of the müllerian ducts must be sought, especially when a disparity between normal puberty and absent menses is encountered.

As will be seen in the discussion of the work-up, the diverse etiologic possibilities for delayed puberty are best classified by the level of gonadotropin encountered. The distribution of diagnostic frequencies in the three categories — hypergonadotropic hypogonadism, hypogonadotropic hypogonadism and eugonadism — are depicted below, representing the findings in 326 patients.

<table>
<thead>
<tr>
<th>Relative Frequency of Delayed Pubertal Abnormalities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypergonadotropic Hypogonadism</strong></td>
</tr>
<tr>
<td>Ovarian failure, abnormal karyotype</td>
</tr>
<tr>
<td>Ovarian failure, normal karyotype</td>
</tr>
<tr>
<td>46, XX</td>
</tr>
<tr>
<td>46, XY</td>
</tr>
<tr>
<td><strong>Hypogonadotropic Hypogonadism</strong></td>
</tr>
<tr>
<td>Reversible</td>
</tr>
<tr>
<td>Physiologic delay</td>
</tr>
<tr>
<td>Weight loss/anorexia</td>
</tr>
<tr>
<td>Primary hypothyroid</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia</td>
</tr>
<tr>
<td>Cushing's syndrome</td>
</tr>
<tr>
<td>Prolactinomas</td>
</tr>
<tr>
<td><strong>Irreversible</strong></td>
</tr>
<tr>
<td>GnRH deficiency</td>
</tr>
<tr>
<td>Hypopituitarism</td>
</tr>
<tr>
<td>Congenital CNS defects</td>
</tr>
</tbody>
</table>
The basic and essential laboratory procedure is a left hand-wrist x-ray for bone age. The Bayley-Pinneau tables predict future adult height, utilizing the bone age and relationship at the proper level of mutual respect. A willingness to listen to problems, together with an adult-to-adult attitude, will place the adolescent-clinician does not like to be different. Therefore, excessive or insufficient growth is not a problem to be dismissed lightly, and psychologic support and reassurance are key features in the management of such problems. In hypogonadotropic delayed puberty, the gonadotropin levels are low, indicating a failure of the hypothalamus to stimulate the pituitary gland. The gonadotropins, LH and FSH, are produced by the pituitary gland and stimulate the ovaries to produce estrogen and progesterone. Inadequate production of these hormones results in the absence of ovulation and menstruation. Poor nutrition (anorexia nervosa, malabsorption, chronic illness, regional ileitis, renal disease) can lead to hypogonadotropic delayed growth and development. Exercise and stress-induced amenorrhea can also delay puberty. Unfortunately, illegal drug use (especially marijuana) must be considered. In the presence of normal ovulation and normal prolactin levels, exclusion of pituitary, parapituitary, or hypothalamic tumor by specialized neuroradiologic procedures is necessary. If tumor or vascular malformation is not found, the diagnosis is (by exclusion) physiological delayed puberty. Craniohypophyseal tumors are usually suprasellar. Other pituitary adenomas, such as adenomas of the pituitary stalk or parapituitary (paraganglioma), can also cause delayed puberty. Hypothalamic amenorrhea, or hypothalamic hypogonadism, is a common cause of delayed puberty. In this condition, the hypothalamus fails to stimulate the pituitary gland to produce gonadotropins. The lack of gonadotropins results in the absence of menstrual cycles. Inappropriate positive feedback (APF) can lead to hyperprolactinemia, which in turn can suppress gonadotropin release and lead to anovulation. Inadequate positive feedback (APF) can result from a deficiency in the feedback loop, which regulates the production of gonadotropins. This can be caused by a deficiency in the production of luteinizing hormone-releasing hormone (LHRH) or a defect in the action of LHRH on the pituitary gland. Decreased secretion of LH (less than 6 IU/L), associated with depressed FSH, is seen in hypothalamic amenorrhea, amenorrhea and anosmia — Kallmann's syndrome, pituitary (tumor) disorders, hyperprolactinemia, or nonpathologic constitutional (physiologic) delay in development. Physiological delayed puberty can be regarded as a physiological variant in development. The typical patient with physiological delay is short with appropriate bone age at diagnosis. Physiological delay accounts for only 10% of cases with delayed puberty, emphasizing the need to seek another diagnosis. As previously noted, physiological delay is frequently seen in a familial pattern with the expectation of a late but otherwise normal growth and pattern and adult reproductive function.

To use the tables, one needs a measurement of height, the patient's age, and an x-ray of the left hand and wrist for bone age. All of the hand and wrist bones should be fully ossified to accurately assess bone age. The tables are used to estimate the patient's adult height by comparing their bone age to their chronological age. If the bone age is greater than the chronological age, the patient is considered to be chronically delayed. If the bone age is less than the chronological age, the patient is considered to be chronically advanced. The tables are not accurate for patients with bone age discrepancies of more than 2 years. Laboratory work-up of delayed puberty usually includes x-rays for bone age, skull imaging (if hypogonadotropic), gonadotropin and prolactin levels, appropriate adrenal and gonadal steroid measurements, and assessment of thyroid function. In addition, general laboratory screening for systemic disorders is worthwhile. Evaluation according to the program outlined in Chapter 11 will lead to the proper diagnosis. Patients with elevated gonadotropins require a karyotype. Measurements of serum IGF-I, IGFBP-3, and IGFBP-2 can distinguish between delayed puberty and growth hormone deficiency; normal levels according to age, body mass index, and stage of puberty are available in the literature.

### Laboratory Assessments of Delayed Puberty

The first priority in therapy is removal or correction of primary etiology when possible. In this regard, thyroid therapy for hypothyroidism, growth hormone for isolated growth hormone deficiency, and treatment of ileitis are examples of specific therapy. In XY individuals, properly timed gonadectomy followed by sex hormone treatment is required. In physiological delay, reassurance that the anticipated development will occur is the only management step needed, especially when there is a family history of delayed puberty. Early hormone treatment is worthwhile in order to minimize psychological stress.

In hypogonadotropic hypogonadism, hormonal therapy will initiate and sustain maturation and function of secondary sexual characteristics and promote the achievement of full height potential. The importance of the adolescent increase in bone density should not be underrated. This is sufficient reason to recommend hormone treatment. Hormone treatment should conform to what we have learned about the early stages of puberty. Very small amounts of estrogen will promote growth and development. Start with unopposed estrogen, 0.3 mg conjugated estrogens or 0.5 mg estradiol daily. After 6 months to 1 year, move to a sequential program with 0.625 mg conjugated estrogens or 1.0 mg estradiol daily and 5 mg medroxyprogesterone acetate or an equivalent progestin for the first 14 days each month. Patients with physiological delay of puberty will continue development on their own when bone age has advanced to 13 years. Monthly menstruation is an important experience for adolescents. Regular and visible bleeding serves to reinforce the young patient's identification with the feminine gender role. However, remember that the doses used for this therapy will not protect against pregnancy in the event the hypothalamic-pituitary-ovarian axis is resistant ovary syndrome. See the discussion in Chapter 11 under "Premature Ovarian Failure."
epiphyses and those of the distal end of the arm are used to determine the skeletal age. The Bayley-Pinneau tables begin at the end of this chapter.

To predict a patient's adult height, use the tables as follows. Go down the left column to the patient's present height, follow this horizontal row to the column under the bone age which is given by 6-month intervals across the top. The number at the intersection represents the predicted adult height. The predicted height can be easily extrapolated if figures do not fall at the 1-inch or 6-month intervals used on the tables.

It is important to use the table suitable for the rate of maturing. If the bone age is within 1 year of the chronologic age, use the table for average girls; if the bone age is delayed 1 year or more, use the table for retarded girls; if the bone age is retarded 1 year or more, use the table for delayed girls. Girls with idiopathic central precocious puberty are an exception to this recommendation. With idiopathic GnRH-dependent precocious puberty, despite advanced bone age, height predictions are more accurate using the Bayley-Pinneau tables for average girls.

The tables are for use with bone age films of the hand and wrist only in conjunction with the Greulich-Pyle Atlas. Use with bone age determined by any other method is less accurate.

Short Stature
Thorough medical history and physical examination will eliminate the usual disorders associated with short stature: malnutrition, chronic urinary tract disease, chronic infectious disease, hypothyroidism, mental illness, panhypopituitarism, and gonadal dysgenesis. In the history, the heights and weights of parents, siblings, and relatives should be obtained along with timing of growth in the family, dietary history, daily activities, and sleep habits. Normal history and examination in an individual with a bone age only 1 year behind the chronologic age suggest a constitutional pattern that does not require treatment.

Endocrine disease is an uncommon basis for impairment of growth. Congenital hypothyroidism is the most frequent problem of this type, followed by hypopituitarism, hypothyroidism with onset during childhood, and excess cortisol.

It is unlikely that a patient with congenital hypothyroidism will present undiagnosed and untreated as an adolescent. However, juvenile hypothyroidism must be suspected in an adolescent with obesity and short stature and normal early childhood development. Similarly, an adolescent with hypopituitarism due to a slow growing pituitary tumor may present with a failure to develop secondary sexual characteristics and a failure to grow. Cortisol excess may be due to Cushings disease (rare in childhood) or to therapy with corticosteroids. Excess endogenous or exogenous corticosteroids suppress skeletal maturation and growth. Moderate overdosage of cortisol; e.g., when treating children with adrenal hyperplasia, can suppress growth.

Treatment of Short Stature. Support and observation are indicated if the physician concludes that an adolescent suffers from a delay of normal growth and no disease process is present. Reassurance, but the family history reveals a consistent pattern of retarded but eventual normal growth. It is helpful to point out the x-ray, indicating that the individual has 1 year or more of unused potential in which to catch up with her friends.

Hormone treatment can be considered when continued failure to grow is evident in the absence of disease. Presently the use of growth hormone is limited to use in growth hormone deficiency. The increment in height achieved with growth hormone treatment of short, but normal, children is too small (about 2.5 cm) to warrant such treatment. Ilicit sources of growth hormone have been administered by parents and young people eager to "grow" to achieve greater athletic prowess. This dangerous practice often leads to growth of fragile bones unsupported by growth hormone.

Anabolic-androgenic steroids are illegally utilized by both adolescent males and females to increase athletic performance and even in an effort to look better. Response to these agents ranges from increased strength and libido (virilization and menstrual dysfunction in women) to liver diseases, impotence, and oligospernia. Excessive androgen use by adolescents can prevent individuals from reaching their genetic height potential. Although not well studied, most experts believe that there are significant psychological and behavioral effects (such as enhanced aggression), as well as psychological and social dependance. In addition, adolescents who use anabolic steroids are more likely to use other drugs and to share needles (a major risk factor for human immunodeficiency virus infection).

Fortunately, it is rare to see a female adolescent complaining of short stature. More commonly it is an adolescent boy who is sensitive to reduced growth, and in whom the use of testosterone may be indicated. In cases of gonadal failure, estrogen can be used in a female to stimulate epiphyseal growth, bringing the bone age to match the chronologic age. Conjugated estrogens (0.3 mg) or estradiol (0.5 mg) administered daily are effective in hypogonadal individuals (this is a much smaller dose than previously used). Patients should be observed at monthly intervals to document the pattern of growth and development. Hormone treatment may be discontinued when the bone age matches the chronologic age.

Tall Stature
This is rarely a problem in boys. Basketball has provided a ready outlet, and fortunately participation in sports is now appealing to girls as well. But girls who are the daughters of very tall parents may come for help. The Bayley-Pinneau tables are accurate in predicting the height of tall girls. A predicted height greater than 6 feet probably deserves treatment.

A hand-wrist x-ray for bone age is necessary. The degree of development of secondary sexual characteristics is important, because the more mature a girl is, the less effective treatment is in influencing her eventual height.

Treatment of Tall Stature. It is difficult to make a decision for treatment, and parental participation in the decision is essential. In a case where some success can be achieved, the patient is relatively young and may find it hard to know what to think about the future problem. Because the adolescent growth spurt precedes menarche, treatment must begin before menarche in order to be optimally successful. This would be as early as 8 or 9 years, and certainly before the age of 12. However, treatment begun after menarche may still achieve up to an inch of growth reduction. Once begun, treatment must continue until epiphyses are fused. If treatment is stopped earlier, further growth will occur. The parents and patient must be informed of possible problems with menstruation, breast symptoms, and water retention.

A simple regimen that provides high doses of estrogen is to use oral contraceptives: a 50 μg ethinyl estradiol oral contraceptive provides 10 times the estrogen usually administered to postmenopausal women. Hand-wrist films should be taken every 6 months until epiphyseal closure is demonstrated. In view of the sensitivity of growth physiology to low levels of estrogen, it is not certain that these high doses are necessary. Although 10-year follow-up of a large number of women treated with high doses of estrogens did not detect any adverse consequences on reproductive function, it would be reasonable to consider the usual replacement dose (0.625–1.25 mg conjugated estrogens or 1–2 mg estradiol), especially if the high doses elicit unpleasant symptoms.

Chapter References
Bayley-Pinneau Table for Accelerated Girls

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.

Bayley-Pinneau Table for Retarded Girls

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.

Copyright ©
Definition of Amenorrhea

Basic Principles in Menstrual Function

Evaluation of Amenorrhea

Step 1

Step 2

Step 3

High Gonadotropins

The Need for Chromosome Evaluation

Premature Ovarian Failure: A Clinical Dilemma

Normal Gonadotropins

Low Gonadotropins

Twinning of the Sella Turcica

Evaluation of the Abnormal Sella Turcica and/or High Prolactin

Hypogonadotropic Hypogonadism

Mullerian Anomalies

Androgen Insensitivity (Testicular Feminization)

Compartment II: Disorders of the Ovary

Turner Syndrome

Mosaicism

XY Gonadal Dysgenesis

Genital Anomalies

The Resistant Ovary Syndrome

Premature Ovarian failure

The Effect of Radiation and Chemotherapy

Compartment III: Disorders of the Anterior Pituitary

Treatment of Nonfunctioning Adenomas

Pituitary Prolactin-Secreting Adenomas

Summary: Therapy of Pituitary Prolactin-Secreting Adenomas

Pregnancy and Prolactin Adenomas

The Empty Sella Syndrome

Sheehan's Syndrome

Compartment IV: Central Nervous System Disorders

Hypothalamic Amenorrhea

Weight Loss, Anorexia, Bulimia

Exercise and Amenorrhea

Eating Disorders and Pregnancy

Inherited Genetic Defects

Postpill Amenorrhea

Hormone Therapy

Chapter References

Compartment I: Disorders of the Endometrium

Premenstrual Syndrome and Premenstrual Dysphoric Disorder

Luteal Phase Deficiency

Premenstrual Tension Syndrome

Premenstrual Dysphoric Disorder

The Endometrium and Menstrual Function

Evaluation of the Abnormal Sella Turcica and/or High Prolactin

The Need for Chromosome Evaluation

Premature Ovarian Failure: A Clinical Dilemma

Normal Gonadotropins

Low Gonadotropins

Twinning of the Sella Turcica

Evaluation of the Abnormal Sella Turcica and/or High Prolactin

Hypogonadotropic Hypogonadism

Mullerian Anomalies

Androgen Insensitivity (Testicular Feminization)

Compartment II: Disorders of the Ovary

Turner Syndrome

Mosaicism

XY Gonadal Dysgenesis

Genital Anomalies

The Resistant Ovary Syndrome

Premature Ovarian failure

The Effect of Radiation and Chemotherapy

Compartment III: Disorders of the Anterior Pituitary

Treatment of Nonfunctioning Adenomas

Pituitary Prolactin-Secreting Adenomas

Summary: Therapy of Pituitary Prolactin-Secreting Adenomas

Pregnancy and Prolactin Adenomas

The Empty Sella Syndrome

Sheehan's Syndrome

Compartment IV: Central Nervous System Disorders

Hypothalamic Amenorrhea

Weight Loss, Anorexia, Bulimia

Exercise and Amenorrhea

Eating Disorders and Pregnancy

Inherited Genetic Defects

Postpill Amenorrhea

Hormone Therapy

Chapter References

Few problems in gynecologic endocrinology are as challenging or taxing to the clinician as amenorrhea. The clinician must be concerned with an array of potential diseases and disorders involving, in many instances, unfamiliar organ systems, some carrying morbidity and even lethal consequences for the patient. Not infrequently, the otherwise confident and experienced clinician dismisses the problem as too complex for a busy practice and refers the patient to a “specialist” in the field. In doing so, the nonavailability of sophisticated laboratory techniques is often cited as necessitating the costly and frequently inconvenient transfer of the patient.

The intent of this chapter is to provide a simple mechanism for the differential diagnosis of amenorrhea of all types and chronology, using procedures available to all clinicians. Strict adherence to this design will unerringly pinpoint the organ system locus of disorder leading to the presenting symptom of amenorrhea. Once this is accomplished, the detailed evidence confirming the diagnosis can be sought and the assistance of appropriate specialists (neurosurgeon, internist, endocrinologist, psychiatrist) confidently chosen. In the end, the patient receives the most reliable diagnosis and therapy at minimum cost and optimum convenience. The majority of patients with amenorrhea have relatively simple problems that can be managed easily by the patients’ primary care clinicians. This happy outcome is also consistent with the modern health care emphasis on cost-effectiveness.

The “workup” to be described is not new. With minor modifications, it has been continuously and successfully applied for several decades. Before presenting the diagnostic workup in detail, it is necessary to provide a definition of amenorrhea, designating the appropriate selection of patients. In addition, a brief review of the physiologic mechanism by which a menstrual flow is produced is presented to clarify the logic of the various steps in the diagnostic procedures.

Definition of Amenorrhea

Any patient fulfilling the following criteria should be evaluated as having the clinical problem of amenorrhea:

1. No period by age 14 in the absence of growth or development of secondary sexual characteristics.
2. No period by age 16 regardless of the presence of normal growth and development with the appearance of secondary sexual characteristics.
3. In a woman who has been menstruating, the absence of periods for a length of time equivalent to a total of at least 3 of the previous cycle intervals or 6 months of amenorrhea.

Having affirmed the traditional criteria, let us now point out that strict adherence to these criteria can result in improper management of individual cases. There is no reason to defer the evaluation of a young girl who presents with the obvious stigmata of Turner syndrome. Similarly, the 14-year-old girl with an absent vagina who is otherwise completely normal should not be told to return in 2 years. A patient deserves a considerate evaluation whenever her anxieties, or those of her parents, bring her to a clinician. Finally, the possibility of pregnancy should always be considered.

Another tradition has been to categorize amenorrhea as primary or secondary in nature. Although these stipulations are inherent in the classic definitions noted above, experience has shown that premature categorization of this sort leads to diagnostic omission in certain instances, and frequently, unnecessary and expensive diagnostic procedures. Because the prescribed workup to be detailed here applies comprehensively to all forms of amenorrhea, the classic definitions are not retained.

Basic Principles in Menstrual Function

The clinical demonstration of menstrual function depends on visible external evidence of the menstrual discharge. This requires an intact outflow tract that connects the internal genital source of flow with the outside. As such, the outflow tract requires patency and continuity of the vaginal orifice, the vaginal canal, and the endocervix with the uterine cavity. The presence of a menstrual flow depends on the existence and development of the endometrium lining the uterine cavity. This tissue is stimulated and regulated by the proper quantity and sequence of the steroid hormones, estrogen and progesterone. The secretion of these hormones
Galactorrhea is considered in greater detail in expansion, or erosion of the sella turcica) can be seen, therefore, with primary hypothyroidism and in patients with elevated GnRH and gonadotropin ecretion due to Constant stimulation by hypothalamic-releasing hormones can result in hypertrophy or hyperplasia of the pituitary. The imaging picture of a tumor (distortion, primary hypothyroidism have always been less than 100 ng/mL. unopposed thyrotropin-releasing hormone (TRH) stimulatory effect on the pituitary cells that secrete prolactin. In our experience, prolactin levels associated with the higher the prolactin levels. The duration of the hypothyroidism is important with regard to the mechanism of the galactorrhea; the longer the duration the higher the incidence of galactorrhea and warranted.

At 6–8 weeks of gestation, 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. This represents the maximal oogonal content of the gonad. From this point in time germ cell content will irreversibly decrease until, some 50 years later, the store of oocytes will be finally exhausted. 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 by birth, and it appears as a fibrous streak.

The basic principles underlying the physiology of menstrual function permit formulation of several discrete compartmental systems on which proper menstruation depends. It is useful to employ a diagnostic evaluation that segregates causes of amenorrhea into the following compartments:

- **Compartment I**: Disorders of the outflow tract or uterine target organ.
- **Compartment II**: Disorders of the ovary.
- **Compartment III**: Disorders of the anterior pituitary.
- **Compartment IV**: Disorders of central nervous system (hypothalamic) factors.

### Evaluation of Amenorrhea

A careful history and physical examination should seek the following: evidence for psychological dysfunction or emotional stress, family history of apparent genetic anomalies, signs of a physical problem with a focus on nutritional status, abnormal growth and development, the presence of a normal reproductive tract, and evidence for CNS disease. A patient with amenorrhea is then exposed to a combined therapeutic and laboratory dissection according to the depicted flow diagrams. Because a significant number of patients with amenorrhea also have galactorrhea (nonpuerperal breast secretion), and there are similarities in the evaluation of these two conditions, the workup as described is appropriate for patients who have amenorrhea, galactorrhea, or both. Galactorrhea is an important clinical physical sign, whether it is spontaneous or present only with careful expression by the examiner, unilateral or bilateral, persistent or intermittent. *Hormonal secretions usually come from multiple duct openings in contrast to pathologic discharge that usually comes from a single duct.* Galactorrhea is considered in greater detail in Chapter 16.

Amenorrhea and galactorrhea need be the sole pertinent initial items of information. Although additional data are undoubtedly available at this time, derived from history and physical examination and evaluation of other endocrine glands such as the thyroid and adrenal, these items should not be utilized for diagnostic purposes until the entire workup is completed. Experience has shown that premature diagnostic bias at this point, while frequently accurate, not uncommonly leads to erroneous judgments as well as inappropriate, costly, and useless testing.

**Step 1**

The initial step in the workup of the amenorrheic patient after excluding pregnancy begins with a measurement of thyroid-stimulating hormone (TSH), a prolactin level, and a progestational challenge. The initial step in the patient presenting with galactorrhea, regardless of menstrual history, also includes TSH and prolactin measurements but adds a coned-down, lateral x-ray view of the sella turcica. The x-ray can be safely omitted in those patients who have galactorrhea, but also have regular, ovulatory menstrual cycles.

Only a few patients presenting with amenorrhea and/or galactorrhea will have hypothyroidism that is not clinically apparent. Although it seems rather extravagant to measure TSH in such a large number of patients for such a small return, because treatment for hypothyroidism is so simple and is rewarded by such a prompt return of ovulatory cycles, and, if galactorrhea is present, by a disappearance of the breast secretions (a slower process that can take several months), TSH measurement is warranted.

The duration of the hypothyroidism is important with regard to the mechanism of the galactorrhea; the longer the duration the higher the incidence of galactorrhea and the higher the prolactin levels. This is thought to be associated with declining hypothalamic content of dopamine with on-going hypothyroidism. This would lead to an unopposed thyrotropin-releasing hormone (TRH) stimulatory effect on the pituitary cells that secrete prolactin. In our experience, prolactin levels associated with primary hypothyroidism have always been less than 100 ng/mL.

Constant stimulation by hypothalamic-releasing hormones can result in hypertrophy or hyperplasia of the pituitary. The imaging picture of a tumor (distortion, expansion, or erosion of the sella turcica) can be seen, therefore, with primary hypothyroidism and in patients with elevated GnRH and gonadotropin ecretion due to
The purpose of the progesterational challenge is to assess the level of endogenous estrogen and the competence of the outflow tract. A course of a progesterational agent totally devoid of estrogenic activity is administered. There are 3 choices: parenteral progesterone in oil (200 mg), the oral administration of micronized progesterone (300 mg), or orally active medroxyprogesterone acetate, 10 mg daily for 5 days. The use of an orally active agent avoids an unpleasant intramuscular injection (although this might be necessary when compliance is a concern). The dose of micronized progesterone is relatively high and should be administered at bedtime to avoid side effects. Other hormonal preparations, such as oral contraceptives, are not appropriate because they do not exert a purely progesterational effect.

Within 2–7 days after the conclusion of progesterational medication, the patient will either bleed or not bleed. If the patient bleeds, a diagnosis of anovulation has been reliably and securely established. The presence of a functional outflow tract and a uterus lined by reactive endometrium sufficiently prepared by endogenous estrogen is confirmed. With this demonstration of the presence of estrogen, minimal function of the ovary, pituitary, and CNS is established. In the absence of galactorrhea, with a normal prolactin level, and a normal TSH, further evaluation is unnecessary.

How much bleeding constitutes a positive withdrawal response? The appearance of only a few blood spots following progesterational medication implies marginal levels of endogenous estrogen. Such patients should be followed closely and periodically reevaluated, because the marginally positive response may progress to a clearly negative response, placing the patient in a new diagnostic category. Bleeding in any amount beyond a few spots is considered a positive withdrawal response.

There are two rare situations associated with a negative withdrawal response, despite the presence of adequate levels of endogenous estrogen. In both situations, the endometrium is decidualized, and, therefore, it will not be shed following the withdrawal of exogenous progestin. The first condition finds the endometrium decidualized in response to high androgen levels, e.g., due to the anovulatory state (with polycystic ovaries). In the second unusual clinical situation, the endometrium is decidualized by high progesterone levels associated with a specific adrenal enzyme deficiency.

All anovulatory patients require therapeutic management, and with this minimal evaluation, therapy can be planned immediately. Because of the short latent period in the progression from normal endometrial tissue to atypia to cancer, clinicians are sensitive to the issue of endometrial cancer. But all too often, the clinician believes that this is a problem limited to older age. The critical feature is the duration of exposure to constant, unopposed estrogen. Therefore, even young women, anovulatory for relatively long periods of time, can develop endometrial cancer. If there is any concern, evaluation of the endometrium (with aspiration curettage) is in order. On the other hand, the latent phase for breast cancer is long, perhaps as long as 20 years. Women who are anovulatory when they are young may have an increased risk of breast cancer when they are postmenopausal. This could reflect exposure to unopposed estrogen, or it could be the consequence of infertility and the absence of the protection against breast cancer that pregnancy early in the reproductive years confers. However, some studies have not observed a link between anovulation and the risk of breast cancer (discussed with full references in Chapter 16).

Minimal therapy of anovulatory women requires the monthly administration of a progesterational agent. An easily remembered program is to prescribe 10 mg medroxyprogesterone acetate daily for the first 10 days of each month. Experience with the endometrium in estrogen therapy programs has established the importance of a time period of at least 10 days to provide adequate protection against the growth-promoting effects of constant estrogen. When reliable contraception is essential, the use of low-dose oral contraceptive pills in the usual cyclic fashion is appropriate. Attempts to demonstrate a relationship between pill use and subsequent postpill amenorrhea have not been successful. Anovulation with amenorrhea or oligomenorrhea should not be viewed as a contraindication to the use of oral contraception.

If, at any time, an anovulatory patient fails to have withdrawal bleeding on a monthly progestin program, this is a sign (providing the patient is not pregnant) that she has moved to the negative withdrawal bleed category, and the remainder of the workup must be pursued. The progesterational challenge will occasionally trigger an ovulation in an anovulatory patient. The tip-off will be a later withdrawal bleed, 14 days after the progesterational challenge.

In the absence of galactorrhea and if the serum prolactin level is normal (less than 20 ng/mL in most laboratories), further evaluation for a pituitary tumor is unnecessary provided the patient has undergone a withdrawal bleed. Random single samples for prolactin are sufficient, because variations in the amplitude of the spikes of secretion and the sleep-related and food-related increases appear to be attenuated in both functional and tumor hyperprolactinemic states. If the prolactin is elevated, imaging evaluation of the sella turcica is essential (as discussed below). At this point in the workup, the following statement is a useful clinical rule of thumb:

A positive withdrawal bleeding response to progesterational medication, the absence of galactorrhea, and a normal prolactin level together effectively rule out the presence of a significant pituitary tumor.

Ectopic production of prolactin is rarely encountered. Increased prolactin secretion should draw attention to the pituitary gland. However, to be complete, we should mention case reports with ectopic secretion associated with pituitary tissue in the pharynx, bronchogenic carcinoma, renal cell carcinoma, a gonadoblastoma, and women with amenorrhea and hyperprolactinemia due to a prolactinoma in the wall of an ovarian dermoid cyst or teratoma. 5 8 9 10 11 12
Patients suspected of having a pituitary tumor, the nature of which is uncertain or puzzling, should have their gonadotropin and a-subunit levels measured. However, tumors of the pituitary are relatively common, and most are not nonsecreting, but, in fact, they arise from gonadotroph cells, and they are active. Patients suspected of having a pituitary tumor, the nature of which is uncertain or puzzling, should have their gonadotropin and a-subunit levels measured.

### Step 2

If the course of progestational medication does not produce withdrawal flow, either the target organ outflow tract is inoperative or preliminary estrogen proliferation of the endometrium has not occurred. Step 2 is designed to clarify this situation. Orally active estrogen is administered in quantity and duration certain to stimulate endometrial proliferation and withdrawal bleeding provided that a completely reactive uterus and patent outflow tract exist. An appropriate dose is 1.25 mg conjugated estrogens or 2 mg estradiol daily for 21 days. The terminal addition of an orally active progestational agent (medroxyprogesterone acetate 10 mg daily for the last 5 days) is necessary to achieve withdrawal. In this way the capacity of Compartment I is challenged by exogenous estrogen. In the absence of withdrawal flow, a validating second course of estrogen is a wise precaution.

As a result of the pharmacologic test of Step 2, the patient with amenorrhea will either bleed or not bleed. If there is no withdrawal flow, the diagnosis of a defect in the Compartment I systems (endometrium, outflow tract) can be made with confidence. If withdrawal bleeding does occur, one can assume that Compartment I systems have normal functional abilities if properly stimulated by estrogen.

From a practical point of view, in a patient with normal external and internal genitalia by pelvic examination, and in the absence of a history of infection or trauma (such as curettage), an abnormality of the outflow tract is unlikely. Outflow tract problems include either destruction of the endometrium, generally the result of an overzealous curettage or the result of an infection, or primary amenorrhea resulting from the discontinuity or disruption of the müllerian tube. Abnormalities in the systems of Compartment I are not commonly encountered, and in the absence of a reason to suspect a problem, Step 2 can be omitted.

### Step 3

With the elucidation of the amenorrheic patient’s inability to provide adequate stimulatory amounts of estrogen, the physiologic mechanisms responsible for the elaboration of this steroid must be tested. In order to produce estrogen, ovaries containing a normal follicular apparatus and sufficient pituitary gonadotropins to stimulate that apparatus are required. Step 3 is designed to determine which of these two crucial components (gonadotropins or follicular activity) is functioning improperly.

This step involves an assay of the level of gonadotropins in the patient. Because Step 2 involved administration of exogenous estrogen, endogenous gonadotropin levels may be artificially and temporarily altered from their true baseline concentrations. Hence, a delay of 2 weeks following Step 2 must ensue before doing Step 3, the gonadotropin assay. One should keep in mind that the midcycle surge of LH is approximately 3 times the baseline level. Therefore, if the patient does not bleed 2 weeks after the blood sample was obtained, a high level can be safely interpreted as abnormal.

Step 3 is designed to determine whether the lack of estrogen is due to a fault in the follicle (Compartment II) or in the CNS-pituitary axis (Compartment III and IV). The result of the gonadotropin assay in the amenorrheic woman who does not bleed following a progestational agent will be abnormally high, abnormally low, or in the normal range.

<table>
<thead>
<tr>
<th>Clinical State</th>
<th>Serum FSH</th>
<th>Serum LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal adult female</td>
<td>5–20 IU/L</td>
<td>5–20 IU/L</td>
</tr>
<tr>
<td>Hypogonadotropic state:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepubertal, hypothalamic, or</td>
<td>Less than 5 IU/L</td>
<td>Less than 5 IU/L</td>
</tr>
<tr>
<td>pituitary dysfunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypergonadotropic state:</td>
<td>Greater than 20 IU/L</td>
<td>Greater than 40 IU/L</td>
</tr>
<tr>
<td>Postmenopausal, castrate, or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ovarian failure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### High Gonadotropins

The clinical implications of elevated gonadotropins are very significant, and, therefore, repeated measurements, several months apart, are worthwhile to document a more than transient state. The association between castrate or postmenopausal levels of gonadotropins and absent ovarian follicles due to accelerated atresia is very reliable, but not totally reliable. There are rare situations in which high gonadotropins can be accompanied by ovaries that contain follicles.

1. On rare occasions, tumors can produce gonadotropins. This situation is usually associated with lung cancer and is so infrequent that, with a normal history and physical examination, routine chest x-ray is not warranted in amenorrheic patients.
2. There have been a handful of reports of a single gonadotropin deficiency. The importance of measuring both FSH and LH can be appreciated because a high level of one and a baseline or undetectable level of the other would reveal this rare condition. The rare cases of a true single gonadotropin deficiency are probably due to homozygous mutations in the gonadotropin genes. A case of hypergonadism due to an LH b-subunit mutation and a woman with primary amenorrhea due to a mutation in the FSH b-subunit have been reported. The mutated b-subunit genes produce alterations in the b-subunits that yield no immunoreactivity or bioactivity. Hence, hypergonadism will be associated with one high and one low gonadotropin level. Heterozygote carriers have a problem of relative infertility. Treatment with exogenous gonadotropins will achieve pregnancy in these rare patients. When a high FSH and a normal or low LH are encountered, an elevated level of a-subunit and the presence of a pituitary mass indicate a gonadotroph adenoma.
3. Elevated gonadotropins can be due to a gonadotropin-secreting pituitary adenoma. However, gonadotropin-secreting adenomas are NOT associated with hypergonadism (amenorrhea), and, therefore, are hard to diagnose. There is no specific symptom or symptom complex associated with hypersecretion of gonadotropins. Thus, these tumors are usually diagnosed because of tumor growth that results in headaches and visual disturbances. Previously it was believed that these tumors were very rare and more common in men. This belief was due to the difficulty in recognizing these adenomas, especially in women. However, tumors of the pituitary are relatively common, and most are not nonsecreting, but, in fact, they arise from gonadotroph cells, and they are active. In addition to secreting FSH and, rarely, LH, these tumors secrete high levels of the a-subunit of the glycopeptide hormones, and sometimes only the a-subunit. Patients suspected of having a pituitary tumor, the nature of which is uncertain or puzzling, should have their gonadotropin and a-subunit levels measured.
During the perimenopausal period, it is normal for FSH levels to begin to rise even before bleeding has ceased.\textsuperscript{11} This is true whether the perimenopausal period is premature at age 25–35 or at the usual time. This increase in FSH is associated with a decrease in inhibin. During the perimenopausal period, the remaining follicles may be viewed as the least sensitive of all follicles because they have remained in place and failed to respond to gonadotropins for many years. The rise in FSH prior to menopause is due to the declining inhibin production by either the less competent ovarian follicles or because the cohort of follicles is reduced in number. Attention must be paid to this situation because a period of elevated levels of FSH can be followed by a pregnancy. An elevated FSH level is not an absolute indicator of infertility. It is not unusual to encounter a pregnancy in a woman after a diagnosis of premature ovarian failure.\textsuperscript{12}

5. In the resistant or insensitive ovary syndrome, the patient with amenorrhea and normal growth and development has elevated gonadotropins, despite the presence of ovarian follicles. In this condition, the ovarian follicles are unresponsive to stimulation compared with premature depletion of follicles in the most common type of premature ovarian failure. This syndrome may be due to absent or defective gonadotropin receptors on the follicles or a postreceptor-signaling defect. Molecular biology studies of patients with premature ovarian failure are discovering rare cases of point mutations; e.g., mutations in the gonadotropin receptor genes that prevent ovarian response.\textsuperscript{32} We anticipate that more and more such cases will be identified. In addition, translocations between regions on X and Y chromosomes that share sequence homology have been reported in patients with secondary amenorrhea and ovarian failure. It is also possible that this is a subtle form of autoimmune ovarian failure. In these cases, laparotomy is the only definitive way to evaluate the ovaries, because follicles are contained deep within the ovary, yielding only to a full thickness biopsy.\textsuperscript{32} Because this condition is very rare, and the chance of achieving pregnancy is probably impossible, even with large doses of exogenous gonadotropins, laparotomy is NOT recommended for every patient with amenorrhea and high gonadotropins.

6. Secondary amenorrhea caused by premature ovarian failure can be due to autoimmune disease.\textsuperscript{3} The ovaries contain normal-appearing primordial follicles, but developing follicles are surrounded by nests of lymphocytes and plasma cells with lymphocytic infiltration of the thecal layer of cells. The exact mechanism of resistance to gonadotropins is not known. Antibodies blocking FSH and LH receptors cannot be demonstrated in immunoglobulin G from women with premature ovarian failure.\textsuperscript{24} Most commonly, evidence of abnormal thyroid function is detected, and, therefore, complete thyroid testing (with antibodies) is necessary in all patients with premature ovarian failure. The extensive polyglandular syndrome (autoimmune polyendocrine syndrome) that includes hypoparathyroidism, adrenal insufficiency, thyroiditis, and moniliasis, is rare; at least one gene mutation has been identified in this autosomal recessive disorder.\textsuperscript{24} In patients with adrenal insufficiency and ovarian failure, antibodies have been detected directed against P450scs, the cholesterol side-chain cleavage enzyme essential for steroidogenesis.\textsuperscript{24} It is believed that antibodies can be directed against any of the vital enzymes involved in steroidogenesis. Other rare conditions associated with premature ovarian failure include myasthenia gravis, idiopathic thrombocytopenic purpura, rheumatoid arthritis, vitiligo, and autoimmune hemolytic anemia. Classically, premature ovarian failure precedes adrenal failure, and, thus, a case can be made for continuing adrenal surveillance.\textsuperscript{24} Frequently, various endocrine disorders will be present among family members. Very rare pregnancies have been reported in women with ovarian failure and autoimmune disease, Ovulation has been restored temporarily with corticosteroid treatment, and at least one patient had a temporary spontaneous return of menstrual ovarian activity.\textsuperscript{32} Because pregnancy is extremely unlikely, consideration should be given to donor oocytes. However, an impressive pregnancy rate has been reported by one group combining suppression with a GnRH agonist, corticosteroid treatment, and induction of ovulation with high doses of exogenous gonadotropins.\textsuperscript{24}

7. Galactosemia is a rare inherited autosomal recessive disorder of galactose metabolism due to a deficiency of galactose-1-phosphate uridyl transferase.\textsuperscript{27} The problem in patients with galactosemia is primarily gonadal; fewer oogonia may be the result of a direct toxic effect of galactose metabolites on germ cell migration to the genital ridge.\textsuperscript{32} Premature ovarian failure is common and usually irreversible.

8. The final rare clinical situation associated with high gonadotropins despite the presence of ovarian follicles is that associated with specific enzymatic deficiencies.\textsuperscript{27} An example is the 17-hydroxylase deficiency (P450scCT) is present in both ovaries and the adrenal gland. A patient with a deficiency of 17-hydroxylase is readily detectable because she would present with absent secondary sexual development (sex steroids cannot be produced due to the enzyme block in the adrenal glands and the ovaries), and hypertension, hypokalemia, and high blood levels of progesterone. A deficiency in the aromatase enzymes is another rare cause of hypergonadotropic amenorrhea and a failure of pubertal development.

The Need for Chromosome Evaluation

All patients under the age of 30 who have been assigned the diagnosis of ovarian failure on the basis of elevated gonadotropins must have a karyotype determination. The presence of mosaicism with a Y chromosome requires excision of the gonadal areas because the presence of any testicular component within the gonad carries with it a significant chance of malignant tumor formation. These are highly malignant secondary tumors from germ cells: gonoblastomas, dysgerminomas, yolk sac tumors, and choriocarcinoma. Approximate 30% of patients with a Y chromosome will not develop signs of virilization. Therefore, even the normal-appearing adult woman with elevated gonadotropin levels must be karyotyped. Even if the karyotype is normal, as an added precaution all patients with premature ovarian failure should have an annual pelvic examination. Such preventive care is also indicated because these patients will be receiving hormone therapy. Over the age of 30, amenorrhea with high gonadotropins is best labeled premature menopause. Genetic evaluation is unnecessary because it is essentially unheard of to have a gonadal tumor appear in these patients after the age of 30. Most of these tumors appear before age 20, but a significant number are detected from age 20 to 30.\textsuperscript{26}

The clinician and patient should give consideration to whether it is worth obtaining an expensive karyotype to seek identification of chromosomal abnormalities that have clinical implications for other family members. Deletions of the X chromosome can be responsible for premature ovarian failure.\textsuperscript{32} Accurate diagnosis of these deletions is not essential for decision making regarding the patient; however, the presence of such abnormalities within a family is associated with infertility due to premature ovarian failure. Having this information can influence the family planning decisions of family members. We recommend that women with premature ovarian failure who are less than 63 inches tall (160 cm) be karyotyped because of the close conjunction of the genes responsible for stature and normal ovarian function.

Because an individual with a mosaic karyotype (e.g., XX/XO) can experience normal pubertal development, menses, and even pregnancy before the onset of a premature menopause, it is appropriate to consider obtaining a karyotype, regardless of menstrual pattern, in an adolescent or young woman less than 60 inches tall. Spontaneous puberty has been reported in 10–20% and spontaneous menstruation in 2–5% of girls with Turner syndrome.\textsuperscript{27}

Premature Ovarian Failure: A Clinical Dilemma

Patients with repeatedly elevated gonadotropin levels can be reliably diagnosed as having ovarian failure and can be considered sterile. In the past this was a diagnosis made with great confidence, and careful explanation was given to the patient indicating that future pregnancy was impossible. However, as many as 10–20% of patients presenting with secondary amenorrhea and elevated gonadotropins (with normal karyotypes), several months later, have demonstrated resumption of normal function.\textsuperscript{26,27} Often, this has been associated with the use of estrogen therapy, suggesting that the estrogen may activate receptor formation on follicles, and the high gonadotropins may, thus, stimulate follicular growth and development. On the other hand, spontaneous recovery of ovarian function may be the reason for apparent success with treatment. In some patients, return of normal ovarian function with pregnancy has occurred in the absence of any treatment. Indeed, careful study of estrogen treatment in women with hypergonadotropic amenorrhea could detect no impact on ovarian folliculogenesis.\textsuperscript{42} Although resumption of normal function is extremely rare, it is now necessary to tell patients who fit into this category that there is a very remote possibility of future pregnancy.
A number of cases of ovarian failure have been reported with autoimmune disorders. Indirect immunofluorescence has been used for the detection of ovarian antibodies, but the results have been variable and of little clinical meaning. Comparing several methodologies currently available, only a small number of women with premature ovarian failure test positively, results are inconsistent, and a need or indication for ovarian antibody testing cannot be established. The search for antibodies to gonadotropin receptors has been equally unrewarding.

A reasonable empiric approach is to perform a few selected blood tests for autoimmune disease. These tests should also be obtained periodically (perhaps every few years) as part of the long-term surveillance against associated autoimmune disorders.

**Calcium**

**Phosphorus**

**Fasting glucose**

**A.M. cortisol**

**Free T₄**

**TSH**

Thyroid antibodies, if thyroid function is abnormal

Complete blood count and sedimentation rate

Total protein, albumin:globulin ratio

Rheumatoid factor

Antinuclear antibody

Testing for adrenocorticotropic hormone (ACTH) reserve (with corticotropin stimulation or metyrapone) is not necessary if clinical appearance and other laboratory tests are normal. Periodic surveillance for adrenal failure is in order because ovarian failure usually precedes adrenal failure. It is not worthwhile to seek the presence of antiovian antibodies in serum or antibodies to the gonadotropins; there is no correlation between the clinical state and the results from such testing.

Other than hypothyroidism, it is uncommon to encounter other disorders associated with premature ovarian failure. In a series of 119 patients with premature ovarian failure and normal karyotypes, recruited by the National Institute of Child Health and Human Development in Bethesda, 10 new cases of hypothyroidism and 3 new cases of diabetes mellitus were discovered; there were no new cases of adrenal insufficiency, hypoparathyroidism, or pernicious anemia.

It has been suggested that blood gonadotropins and estradiol should be measured weekly on 4 occasions. If FSH is not higher than LH (an FSH:LH ratio less than 1.0), and if estradiol is greater than 50 pg/mL, induction of ovulation can be considered. Prior to treatment with exogenous gonadotropins, no advantage has been demonstrated to first bringing the elevated gonadotropins down to normal range with administration of a GnRH agonist. The experience with this approach has been unrewarding, and we do not recommend it.

The clinician needs to compare the choice of a full-thickness ovarian biopsy with empirical treatment. In our view, empirical treatment appears to outweigh the cost and risk of laparotomy, since only a very rare woman with hypergonadotropic amenorrhea can be expected to conceive. In other words, even if there is some gonadotropic and estradiol evidence of follicular activity, the response to exogenous gonadotropin stimulation has been very disappointing. We believe that ovarian biopsy is not indicated or necessary.

Transvaginal ultrasonography can identify ovarian follicular activity in many of these patients. However, the meaning of this finding is uncertain. Are patients with ovarian follicles present on ultrasonography more likely to respond to treatment and achieve pregnancy? In time, clinical studies should provide us with an answer.

This clinical problem requires scientific study. Ovarian failure may be the consequence of more than one abnormal condition. One recognized example is ovarian failure due to an accelerated rate of follicular atresia leading to the premature depletion of the follicular supply. Other cases may be due to a failure in the gonadotropin receptor mechanism. Unsuccessful binding can be due to defective receptors, blockage of receptors (e.g., by autoantibodies), or to genetic mutations that direct the production of gonadotropins of altered structure which are incapable of binding. Molecular studies of these patients should bring important clinical clarification in the coming years.

**Normal Gonadotropins**

**Low Gonadotropins**

If the gonadotropin assay is abnormally low, or in the normal range, one final localization is required to distinguish between a pituitary (Compartment III) or CNS-hypothalamic (Compartment IV) cause for the amenorrhea. This is achieved by imaging evaluation of the sella turcica for signs of abnormal change.

**Imaging of the Sella Turcica**

The diagnostic modality of choice is either thin-section coronal computed tomography (CT scan) with intravenous contrast enhancement or magnetic resonance imaging (MRI) with gadolinium enhancement. CT scanning (capable of high-resolution 1.5 mm cuts) is able to evaluate the contents of the sella turcica as well as the suprasellar area, however, total accuracy is not achieved. Magnetic resonance imaging is even more sensitive than the CT scan, but it is also more expensive, and it requires a lengthy period of time to obtain the images. MRI provides highly accurate assessments without biologic hazard, and it is better for evaluation of extrascellar extensions and the empty sella turcica. Most neuroradiologists and neurosurgeons prefer MRI.
Microadenomas are very common. There is a significant recurrence rate after surgery. There is no contraindication to hormone therapy or oral contraception.

Frankly, the endocrine maneuvers yield no more useful information than the two major screening procedures, the blood prolactin and the coned-down view of the sella turcica rule. Most consistently abnormal response (a blunter response of prolactin), but some patients with tumors respond normally. Variability in response to all maneuvers is the include GnRH stimulation, TRH stimulation, and other steps to alter prolactin, growth hormone, and ACTH secretion. TRH stimulation of the prolactin response is the most consistently abnormal response (a blunter response of prolactin), but some patients with tumors respond normally. Variability in response to all maneuvers is the rule. Frankly, the endocrine maneuvers yield no more useful information than the two major screening procedures, the blood prolactin and the coned-down view of the sella turcica.

The presence of visual problems and/or headaches should also encourage CT scan or MRI evaluation. Headaches are definitely correlated with the presence of a pituitary adenoma. Although they are usually bifrontal, retro-orbital, or bitemporal, no locations or features are specific for pituitary tumors.

The prolactin level of 100 ng/mL for determining a more aggressive approach has been empirically chosen. Both in our own experience, and that of others, large tumors are most frequently associated with prolactin levels greater than 100 ng/mL. Large masses associated with prolactin levels less than 100 ng/mL are more likely to be tumors other than prolactin-secreting adenomas, causing stalk compression and interruption of the normal dopamine regulation of prolactin secretion. These tumors will be associated with abnormal changes present in the coned-down view of the sella turcica.

If imaging rules out an empty sella syndrome, or a suprasellar problem, treatment is dictated by the patient's desires, the size of the tumor, and the rapidity of growth of the tumor.

The above approach to the problem of pituitary tumors implies that patients with prolactin levels less than 100 ng/mL and with normal coned-down views of the sella turcica can be offered a choice between treatment and surveillance. An annual prolactin level and a periodic coned-down view (at first annually, and then at increasing intervals) are indicated for continued observation to detect an emerging and slow-growing tumor. Dopamine agonist therapy is recommended for patients wishing to achieve pregnancy and for those patients who have galactorrhea to the point of discomfort. Thus far, long-term therapy with a dopamine agonist has not been proven to be successful in producing a complete reversal of the problem (with either permanent suppression of elevated prolactin levels or elimination of small tumors). Thus, a very strong argument can be made for a "need not to know" the presence of a pituitary microadenoma. If treatment and management are not changed, it is not necessary to document the presence of a microadenoma.

Reasons Why the Diagnosis of Microadenoma Is Not Necessary

1. Microadenomas are very common.
2. Microadenomas very rarely grow during pregnancy.
3. Microadenomas very rarely progress to a macroadenoma (10 mm or more in diameter).
4. There is a significant recurrence rate after surgery.
5. The natural course is unaffected by dopamine agonist treatment.
6. There is no contraindication to hormone therapy or oral contraception.
7. It is better to avoid the problem of the pituitary incidentaloma.

Contemporary reviews point out the short comings of the coned-down view of the sella turcica, citing the limitation of excluding only macroadenomas. Indeed, reviews of our text have faulted this chapter for being less than the state-of-the-art, emphasizing that where we use the coned-down view, MRI should be uniformly obtained. We would argue that the state-of-the-art approach is to use the MRI when necessary, to avoid the compulsion to document the presence of a microadenoma for the reasons stated above, to be cost-effective. This takes strength of conviction when your radiologist reports that a coned-down view of the sella turcica is not sufficient.

The Pituitary Incidentaloma

The percentage of pituitary glands found to contain unsuspected adenomas, all microadenomas, ranged from 9% to 27% in autopsy series. Therefore, many individuals, probably 10%, have silent pituitary masses that are endocrinologically inactive and have no adverse effects on well-being. These lesions are being found incidentally because of the increased use of CT scanning and MRI. Silent microadenomas (less than 10 mm diameter) do not grow, and even macroadenomas (10 mm or more diameter) grow slowly and only rarely. This benign course argues against immediate intervention in patients who have no evidence of hormonal disturbances; long-term surveillance is appropriate. Imaging reassessment of a microadenoma should be obtained at 1, 2, and 5 years, and if there is no change, no further studies are necessary. Imaging reassessment of a macroadenoma should be obtained at 0.5, 1, 2, and 5 years. Growth, of course, requires treatment. Although it is reasonable to screen hormonal function in the presence of a macroadenoma, the value of hormonal screening in a normal individual with a microadenoma is debatable. Minimal hormonal screening includes prolactin, TSH, IGF-I, and a 24-hour urinary cortisol or the overnight dexamethasone suppression test (Chapter 13).

Evaluation of the Abnormal Sella Turcica and/or High Prolactin

The high incidence of pituitary tumors in patients with amenorrhea has prompted a search for a reliable method of diagnosing the condition. Expectations for the utilization of endocrine testing to discriminate between disorders of the hypothalamus and the anterior pituitary have not been realized. These endocrine maneuvers include GnRH stimulation, TRH stimulation, and other steps to alter prolactin, growth hormone, and ACTH secretion. TRH stimulation of the prolactin response is the most consistently abnormal response (a blunter response of prolactin), but some patients with tumors respond normally. Variability in response to all maneuvers is the rule.

Frankly, the endocrine maneuvers yield no more useful information than the two major screening procedures, the blood prolactin and the coned-down view of the sella
turbica. Visual field examination is not useful in screening for pituitary tumors because abnormalities are seen only with large tumors that are evident by prolactin, x-ray evaluation, and/or visual symptoms and headaches.

If the coned-down view is abnormal and/or the prolactin level is over 100 ng/mL, further evaluation and treatment require consultation with expert endocrine resources. These patients are rare, and accumulated experience that can provide the necessary clinical judgment can be found only with the referral resource. On the other hand, our workup easily deals with the vast majority of patients, and the few who require a multidisciplinary team approach are readily identified.

Hypogonadotropic Hypogonadism

Patients with amenorrhea and without galactorrhea who have reached this point in the workup and have normal imaging studies are classified as hypothalamic amenorrhea. The mechanism of the amenorrhea is suppression of pulsatile GnRH secretion below its critical range. This is a diagnosis by exclusion because we can identify probable causes (e.g., anorexia and weight loss) but we cannot test, manipulate, or measure the hypothalamus to prove our diagnosis.

Specific Disorders Within Compartments

With only modest effort, expense, and time, the problem of amenorrhea has been dissected into compartments of dysfunction which positively correlate with specific organ systems. At this point, with the specific anatomic locus of the defect defined, the clinician can now undertake steps to elucidate the specific disorder leading to amenorrhea. Congenital abnormalities are limited to amenorrhea that presents in the pubertal period of life. In a collection of 262 patients with secondary amenorrhea of adult onset, the following diagnostic frequencies were most often observed:

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Condition</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Asherman's syndrome</td>
<td>7.0%</td>
</tr>
<tr>
<td>II</td>
<td>Abnormal chromosomes</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>Normal chromosomes</td>
<td>10.0%</td>
</tr>
<tr>
<td>III</td>
<td>Prolactin tumors</td>
<td>7.5%</td>
</tr>
<tr>
<td>IV</td>
<td>Anovulation</td>
<td>28.0%</td>
</tr>
<tr>
<td></td>
<td>Weight loss/anorexia</td>
<td>10.0%</td>
</tr>
<tr>
<td></td>
<td>Hypothalamic suppression</td>
<td>10.0%</td>
</tr>
<tr>
<td></td>
<td>Hypothyroidism</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Compartment I: Disorders of the Outflow Tract or Uterus

Asherman's Syndrome

Secondary amenorrhea follows destruction of the endometrium (Asherman's syndrome). This condition generally is the result of an overzealous postpartum curettage resulting in intrauterine scarification. A typical pattern of multiple synchieae is seen on a hysterosgram. Diagnosis by hysteroscopy is more accurate and will detect minimal adhesions that are not apparent on a hysterosgram. In the presence of normal ovarian function, the basal body temperature will be biphasic. The adhesions may partially or completely obliterate the endometrial cavity, the internal cervical os, the cervical canal, or combinations of these areas. Surprisingly, despite stenosis or atresia of the internal os, hematometra does not inevitably occur. The endometrium, perhaps in response to a buildup of pressure, becomes refractory, and simple cervical dilatation cures the problem. Asherman's syndrome also can occur following uterine surgery, including cesarean section, myomectomy, or metotropath. Very severe adhesions have been noted following postpartum curettage and postpartum hypogonadism; e.g., in Sheehan's syndrome.

Patients with Asherman's syndrome can present with other problems besides amenorrhea, including miscarriages, dysmenorrhea, or hypomenorrhea. They can even have normal menses. Infertility can be present with mild adhesions, an association not readily explainable. Patients with repeated miscarriages, infertility, or pregnancy wastage should have investigation of the endometrial cavity by hysterosgram or hysteroscopy.

Impairment of the endometrium resulting in amenorrhea can be caused by tuberculosis, a condition that is rare in the United States. Diagnosis is made by culture of the menstrual discharge or tissue obtained by endometrial biopsy. Uterine schistosomiasis is another rare cause of end organ failure, and eggs of the parasite may be found in urine, feces, rectal scrapings, menstrual discharge, or endometrium. We have seen the syndrome following intrauterine device (IUD)-related infections and severe, generalized pelvic infections.

Asherman's syndrome in the past was treated with a dilatation and curettage to break up the synchieae and, if necessary, an on-the-table hysterosgram to ensure a free uterine cavity. Hysteroscopy with direct lysis of adhesions by cutting, cautery, or laser yields better results than the "blind" dilatation and curettage. Following operation, a method should be utilized to prevent the sides of the uterine cavity from adhering. Previously an IUD was used for this purpose; however, a pediatric Foley catheter appears to be a better option. The bag is filled with 3 mL fluid, and the catheter is removed after 7 days. A broad-spectrum antibiotic is started preoperatively and maintained for 10 days. An inhibitor of prostaglandin synthesis can be used if uterine cramping is a problem. The patient is treated for 2 months with high stimulatory doses of estrogen (e.g., conjugated estrogens 2.5 mg daily 3 of 4 days with medroxyprogesterone acetate 10 mg daily added during the 3rd week). When the initial attempt fails to re-establish menstrual flow, repeated attempts are worthwhile. Persistent treatment with repeated procedures may be necessary to regain reproductive potential. Approximately 70–80% of patients with this condition have achieved a successful pregnancy. Pregnancy is frequently complicated, however, by premature labor, placenta accreta, placenta previa, and/or postpartum hemorrhage.

Müllerian Anomalies

In primary amenorrhea, discontinuity by segmental disruptions of the müllerian tube should be ruled out. Thus, imperforate hymen, obliteration of the vaginal orifice, and lapses in continuity of the vaginal canal must be ruled out by direct observation. The cervix or the entire uterus may be absent. Far less common, the uterus may be present, but the cavity absent, or, in the presence of a cavity, the endometrium may be congenitally lacking. With the exception of the latter abnormalities, the clinical problem of müllerian anomalies due to obstruction is compounded by the painful distension of hematocolpos, hematometra, or hematoperitoneum. In all instances, an effort must be made to incise and drain from below at the points of closure of the müllerian tube. Even in complicated circumstances reestablishment of müllerian duct continuity usually can be achieved surgically. The unfortunate consequences of operative extirpation of painful masses from above with damage to bladder, ureter, and rectum, as well as irretrievable loss of distended but otherwise healthy reproductive organs, are rare but well remembered.

Knowing what to expect prior to attempting surgical correction is a great advantage. Imaging (MRI) can be utilized to accurately delineate the anatomical abnormality. A correct preoperative diagnosis will certainly facilitate the planning and execution of surgery.

Müllerian Agenesis

Lack of müllerian development (Mayer-Rokitansky-Kuster-Hauser syndrome) is the diagnosis for the individual with primary amenorrhea and no apparent vagina. This is a relatively common cause (about 1 in 4000 female births) of primary amenorrhea, more frequent than congenital androgen insensitivity and second only to gonadal dysgenesis. These patients are an absence or hypoplasia of the internal vagina, and usually an absence of the uterus and fallopian tubes. However, rarely, the uterus may be normal, but lacking a conduit to the introitus, or there may only be rudimentary, bicornuate cords present. If a partial endometrial cavity is present, this condition is considered müllerian agenesis.

The exact cause of müllerian agenesis is unknown; however, likely causes are mutations of the gene for antimitlerian hormone or the gene for the antimüllerian
hormone receptor. The underlying mechanism would be unwanted exposure to antimüllerian hormone activity. Thus far, no activating mutations have been reported, in contrast to inactivating mutations that cause persistence of müllerian structures. Although usually sporadic, occasional occurrence may be noted within a family. A mutation has been identified in galactose-1-phosphate uridyl transferase in daughters with müllerian agenesis and their mothers. This is different from classic galactosmia, however, it is postulated that increased intrauterine exposure to galactose because of this error in galactose metabolism can be a biologic basis for müllerian agenesis. High-galactose feeding of pregnant mice delays vaginal opening in female offspring. In this group of patients with müllerian agenesis, oocyte depletion (and premature ovarian failure) may be more common.

Further evaluation should include radiologic studies. Approximately one-third of patients have urinary tract abnormalities, and 12% or more have skeletal anomalies, most involving the spine, although absent digits and syndactyly (webbing or fusion of fingers or toes) can occur. Renal tract abnormalities include ectopic kidney, renal agenesis, horseshoe kidney, and abnormal collecting ducts. When the presence of a uterine structure is suspected on examination, ultrasound can be utilized to depict the size and symmetry of the structure. When the anatomic picture on ultrasonography is not certain, MRI is indicated. Laparoscopic visualization of the pelvis is not necessary. MRI is more accurate than ultrasonography and less expensive and invasive than laparoscopy. Explantation of the müllerian remnants is certainly not necessary unless they are causing a problem such as uterine fibroid growth, hematometra, endometriosis, or symportic herniation into the inguinal canal.

Because of the difficulties and complications experienced in surgical series, we favor, when possible, an alternative to the surgical construction of an artificial vagina. Instead, we encourage the use of progressive dilatation as initially described by Frank and later by Wabrek et al. Beginning first in a posterior direction, and then after 2 weeks changing upward to the usual line of the vaginal axis, pressure with commercially available vaginal dilators is performed for 20 min daily to the point of modest discomfort. Utilizing increasingly larger dilators, a functional vagina can be created in several months. Plastic syringe covers can be used instead of the expensive commercial glass dilators. An easier and very effective technique is to hold the dilator in place with a tight garment, maintaining pressure by sitting on a racing bicycle seat (mounted on a special stool or even on a bicycle).

In patients who are unwilling or unable to undergo the dilatation program, the Vecchietti operation applies a traction device either transabdominally or by laparoscopy. Postoperative traction creates a functional vagina in 7–9 days.

Operative treatment should be reserved for those women in whom the Frank method is unacceptable, or fails, or when a well-formed uterus is present and fertility might be preserved. The symptoms of retained menstruation should identify these patients. One recommendation is to perform an initial laparotomy to evaluate the cervical canal; if the cervix is atretic, the uterus should be removed. If it is the relatively simple problem of an imperforate hymen or a transverse vaginal septum, surgery is indicated. Most have recommended against trying to preserve fertility in the presence of complete vaginal agenesis. The morbidity subsequent to this surgery argues for removal of the müllerian structures at the time of construction of a neovagina.

Patients with a transverse vaginal septum, which is a failure of canalization of the distal third of the vagina, usually present with symptoms of obstruction and urinary frequency. A transverse septum can be differentiated from an imperforate hymen by a lack of distension at the introitus with Valsalva's maneuver. A transverse vaginal septum can be accompanied by abnormalities of the upper reproductive tract; e.g., absent segments or atresia of the fallopian tubes or unilateral absence of the fallopian tube and ovary.

Distal obstruction of the genital tract is the only condition in this category that can be considered an emergency. Delay in surgical treatment can lead to infertility due to inflammatory changes and endometriosis. Definitive surgery should be accomplished as soon as possible. Diagnostic needling should be avoided because a hematocolpos can be converted into a pyocolpos.

Reassurance and support are necessary to carry a patient through these procedures. Problems with body image and sexual enjoyment can be avoided, and, although infertile, a full and normal life as a woman can be achieved. Furthermore, genetic offspring can be achieved by collection of oocytes from the genetic mother, fertilization by the genetic father, and placement into a surrogate carrier. An analysis of 34 surrogacy live births resulting from oocytes retrieved from 58 women with congenital absence of the uterus and vagina could find no indication of inheritance in a dominant fashion, making surrogate pregnancy a reasonable option for patients with this disorder.

Androgen Insensitivity (Testicular Feminization)

Complete androgen insensitivity (testicular feminization) is the likely diagnosis when a blind vaginal canal is encountered and the uterus is absent (also discussed in Chapter 9). This is the third most common cause of primary amenorrhea after gonadal dysgenesis and müllerian agenesis. The patient with testicular feminization is a male pseudohermaphrodite. The adjective male refers to the gonadal sex; thus, the individual has testes and an XY karyotype. Pseudohermaphroditism means that the genitilia are opposite of the gonads; thus, the individual is phenotypically female but with absent or meager pubic and axillary hair.

<table>
<thead>
<tr>
<th>Differences between Müllerian Agenesis and Testicular Feminization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Müllerian Agenesis</td>
</tr>
<tr>
<td>Karyotype</td>
</tr>
<tr>
<td>Heredity</td>
</tr>
<tr>
<td>Sexual hair</td>
</tr>
<tr>
<td>Testosterone level</td>
</tr>
<tr>
<td>Other anomalies</td>
</tr>
<tr>
<td>Gonadal neoplasia</td>
</tr>
</tbody>
</table>

The male pseudohermaphrodite is a genetic and gonadal male with failure of virilization. Failures in male development can be considered a spectrum with incomplete forms of androgen insensitivity being represented by some androgen response. Transmission of this disorder is by means of an X-linked recessive gene that is responsible for the androgen intracellular receptor (see Chapter 9 for a discussion of the androgen receptor defect). Clinically, the diagnosis should be considered in:

1. A female child with inguinal hernias because the testes are frequently partially descended.
2. A patient with primary amenorrhea and an absent uterus.
3. A patient with absent body hair.

These patients appear normal at birth except for the possible presence of an inguinal hernia, and most patients are not seen by a physician until puberty. Growth and development are normal, although overall height is usually greater than average, and there may be an eunuchoid tendency (long arms, big hands, and big feet). The breasts, although large, are abnormal; actual glandular tissue is not abundant, nipples are small, and the areolae are pale. More than 50% have an inguinal hernia, and later by Wabrek et al. Therefore, once full development is attained after puberty, the gonads should be removed at approximately age 16–18, and the patient should receive hormone therapy. This is the only exception to the rule that gonads with a Y chromosome should be removed as soon as a diagnosis is made. There are two reasons: first, the development achieved with hormone treatment does not seem to match the smooth pubertal changes due to endogenous hormones, and second, gonadal tumors in these patients have not been encountered prior to puberty. Removal of gonadal tissue can be accomplished by a skilled operator through the laparoscope, reserving the option of laparotomy if the gonads are inaccessible.

When testicular feminization was first studied, it was found that the urinary 17-ketosteroids were normal, and it was suggested that there might be a resistance to
androgen action rather than an absence of androgens — a congenital androgen insensitivity. Indeed, the plasma levels of testosterone are in the normal to high male range, and the plasma clearance and metabolism of testosterone are normal. Thus, these patients produce testosterone, but they do not respond to androgens, either their own or those given locally or systemically. Therefore, the critical steps in sexual differentiation, which require androgens, fail to take place, and development is totally female. Because antimüllerian hormone is present, development of the müllerian duct is inhibited, hence the absence of uterus, tubes, and upper vagina.

This syndrome is marked by a unique combination:
1. Normal female phenotype.
2. Normal male karyotype, 46,XY.
3. Normal or slightly elevated male blood testosterone levels and a high LH.

Cases of incomplete androgen insensitivity (one-tenth as common as the complete syndrome) represent individuals with some androgen effect. These individuals may have clitoral enlargement, or a phallus may even be present. Axillary hair and pubic hair develop along with breast growth. Gonadectomy should be deferred in such cases because it will obviate unwanted further virilization. Patients with a deficit in testicular 17b-hydroxysteroid dehydrogenase activity will have impaired testosterone production and present clinically as incomplete androgen insensitivity. Because treatment (gonadectomy) is the same, precise diagnosis is not essential.

In the past, conventional wisdom warned against unthinking and "needless" disclosure of the gonadal and chromosomal sex to a patient with complete androgen insensitivity. This attitude has changed as, more and more, patients desire and appreciate a full understanding of themselves. Although infertile, these patients are certainly completely female in their gender identity, and this should be reinforced rather than challenged. We now strongly advocate combining a truthful education with appropriate psychological counseling of patient and parents. A good resource is the international support group, A1PS, that is headquartered in the United Kingdom (2 Shirburn Avenue, Mansfield, Nottinghamshire, NG18 2BY, UK): [http://www.medhelp.org/www/ais](http://www.medhelp.org/www/ais).

### Compartment II: Disorders of the Ovary

Problems in gonadal development can present with either primary or secondary amenorrhea. From 30 to 40% of primary amenorrhea cases have gonadal streaks due to abnormal development: gonadal dysgenesis. These patients can be grouped according to the following karyotypes:

- 50% — 45, X
- 25% — Mosaics
- 25% — 46, XX

Women with gonadal dysgenesis can also present with secondary amenorrhea. The karyotypes associated with this presentation are, in order of decreasing frequency:

- 46,XX (most common).
- Mosaics (e.g., 45,X/46,XX).
- Deletions in X short and long arms.
- 47,XXX
- 45,X

Both X chromosomes must be present and active in oocytes to avoid the accelerated loss of follicles. The finding of a normal karyotype in patients with ovarian failure, the most common situation, is most perplexing, suggesting subtle reasons for loss of activity, probably due to specific gene alterations. There is reason to believe that specific genes confined to a portion of the X chromosome are necessary for normal ovarian function.

Gonadal dysgenesis associated with a normal karyotype is also linked to neurosensory deafness (Perrault syndrome). Auditory evaluation should be considered in all 46,XX gonadal dysgenesis cases.

Pure gonadal dysgenesis indicates the presence of bilateral streak gonads, regardless of karyotype. Mixed gonadal dysgenesis indicates testicular tissue on one side and a streak gonad on the other.

**Turner Syndrome**

Turner syndrome (an abnormality in or an absence of one of the X chromosomes) is a well known and thoroughly studied entity. (See [Chapter 9](chapter9) for a full discussion.) The characteristics of short stature, webbed neck, shield chest, and increased carrying angle at the elbow, combined with hypergonadotrophic hypogonadotrophic amenorrhea, make a diagnosis possible on the most superficial evaluation. Due to a lack of ovarian follicles, there is no gonadal sex hormone production at puberty, and thus patients present with primary amenorrhea. However, special attention must be given to the less common variations of this syndrome. Autoimmune disorders, cardiovascular abnormalities, and various renal anomalies must be ruled out. A karyotype should be performed on all patients with elevated gonadotropins, despite the appearance of a typical case of Turner syndrome. The presence of a pure syndrome, 45,X chromosome single-cell line, should be confirmed. This expensive test cannot be viewed just as a step toward academic perfection. Forty percent of individuals who appear to have Turner syndrome are mosaics or have structural aberrations in the X or Y chromosome.

**Mosaicism**

The presence of mosaicism (multiple cell lines of varying sex chromosome composition) must be ruled out for a very important reason. The presence of a Y chromosome in the karyotype requires excision of the gonadal areas because the presence of any medullary (testicular) component within the gonad is a predisposing factor to tumor formation and to heterosexual development (virilization). Only in the patient with the complete form of androgen insensitivity can laparotomy be deferred until after puberty, because the individual is resistant to androgens and gonadal tumors occur late. In all other patients with a Y chromosome, gonadectomy should be performed as soon as the diagnosis is made to avoid virilization and early tumor formation. One should be aware that approximately 30% of patients with a Y chromosome will not develop signs of virilization. Therefore, even the normal-appearing adult patient with elevated serum levels of gonadotropins must be karyotyped to detect a silent Y chromosome so that prophylactic gonadectomy can be performed before neoplastic changes occur. The fully stained and banded karyotype continues to be the best method to detect the presence of testicular tissue or other mosaic combinations. When standard cytogenetic analysis is uncertain, expert consultation should be obtained to pursue further analysis with X- and Y-specific DNA probes. Probing for Y chromosome material is also indicated when virilization occurs, despite no apparent Y on the karyotype and when a chromosomal fragment of uncertain origin is identified (discussed in [Chapter 9](chapter9)).

The impact of mosaicism, even in the absence of a Y-containing line, is significant. With an XX component (e.g., XXXO), functional cortical (ovarian) tissue can be found within the gonad, leading to a variety of responses, including some degree of female development, and, on occasion, even menses and reproduction. These individuals may appear normal, attaining normal stature before premature menopause is experienced. More commonly, these patients are short. Most patients with missing sex chromosome material are less than 63 inches (160 cm) in height. The menopause is early, because the functioning follicles undergo an accelerated rate of atresia.

This complex array of gonadal dysgenesis variations, from the typical pure form to an otherwise normal-appearing and functioning woman with premature menopause, is the result of a variety of mosaicism, which produces a complex mixture of cortical and medullary gonadal tissue. The clinical importance of this information justifies obtaining karyotypes in all cases of elevated gonadotropins in women under age 30. Although more uncommon, even autosomal abnormalities can be associated with hypergonadotrophic ovarian failure, e.g., a 28-year-old woman with secondary amenorrhea and elevated gonadotropins has been reported with trisomy 18 mosaicism. All patients with absent ovarian function and quantitative alterations in the sex chromosomes are categorized as having gonadal dysgenesis (Chapter 9).

**XY Gonadal Dysgenesis**

A female patient with an XY karyotype who has a palpable müllerian system, normal female testosterone levels, and lack of sexual development has Swyer's syndrome.
syndrome. Tumor transformation in the gonadal ridge can occur at any age, and extirpation of the gonadal streaks should be performed as soon as the diagnosis is made.

Gonadal Agenesis

No complicated clinical problems accompany the gonadal failure due to agenesis. Without precise information, only conjecture about the causes of absent development can be made. Thus, viral and metabolic influences in early gestation or undiscovered genetic mutations are suspected. Nevertheless, the final result is irretrievable — hypergonadotropic hypogonadism. In the absence of gonadal function, development is female. Surgical removal of the gonadal streaks is necessary to avoid the possibility of neoplasia.

The Resistant Ovary Syndrome

There is a rare patient with amenorrhea and normal growth and development who has elevated gonadotropins despite the presence of unstimulated ovarian follicles, and there is no evidence of autoimmune disease. Laparotomy is necessary to arrive at a correct diagnosis by obtaining adequate histological evaluation of the ovaries. This can demonstrate not only the presence of follicles but the absence of the lymphocytic infiltration seen with autoimmune disease. Because of the rarity of this condition and the very low chance of achieving pregnancy even with high doses of exogenous gonadotropins, we do not believe it is worthwhile to perform a laparotomy for the purpose of ovarian biopsy on every patient with amenorrhea, high gonadotropins, and a normal karyotype. These patients are excellent candidates for oocyte donation.

Premature Ovarian Failure

Premature ovarian failure (the early depletion of ovarian follicles) is surprisingly common. Approximately 1% of women will experience ovarian failure before the age of 40, and in women with primary amenorrhea, the prevalence ranges from 10% to 28%. The etiology of premature ovarian failure is unknown in most cases. It is useful to explain to the patient that it is probably a genetic disorder with an increased rate of follicle disappearance. Often, specific sex chromosome anomalies can be identified. The most common abnormalities are 45,X and 47,XXX, followed by mosaicism and specific structural abnormalities on the sex chromosomes. Searching for 45,X/46,XX mosaicism using fluorescence in situ hybridization, a higher percentage of cells containing a single X chromosome can be detected in women who present with premature ovarian failure. The cause of ovarian failure is most likely accelerated follicular atresia because even 45,X (Turner syndrome) patients begin with a full complement of germ cells. In addition, premature ovarian failure can be due to an autoimmune process, or to destruction of follicles by infections, such as mumps oophoritis, or a physical insult, such as irradiation or chemotherapy.

The problem can present at varying ages, depending on the number of follicles left. It is useful to view the various presentations as representing a stage in the process of perimenopausal change, no matter what the chronological age of the patient. If loss of follicles has been rapid, then primary amenorrhea and lack of sexual development will be present. If loss of follicles takes place during or after puberty, then the extent of adult phenotypic development and the time of onset of secondary amenorrhea will vary accordingly.

In view of the many case reports documenting resumption of normal function in patients with normal karyotypes (10-20% of these patients), we cannot be certain that these patients will be sterile forever. On the other hand, laparotomy and full-thickness ovarian biopsy surely are not necessary for all of these patients. We believe that a minimal approach, with a survey for autoimmune disease (recognizing that there is no practical clinical method to accurately diagnose autoimmune ovarian failure) and an assessment of ovarian-pituitary activity, is sufficient. As with other hypogonadal patients, hormone therapy is recommended. However, because of spontaneous ovulations, an oral contraceptive is a better treatment regimen choice if pregnancy is not desired. The best prospect for pregnancy is with donated oocytes; however, it should be noted that pregnancy rates are reduced when using a sibling's donated oocytes.

Molecular Explanations for Ovarian Failure

A group of patients with ovarian failure and normal chromosomes has been identified in Finland, displaying a recessive inheritance pattern. A point mutation was detected in the FSH receptor gene and demonstrated to be the cause of ovarian failure in this population. This mutation accounted for 29% of 75 Finnish women with ovarian failure. Ovarian follicles were present in these women, as detected by ultrasonography, although the ovaries were small in volume. Finland is known to harbor an increased number of inherited conditions. A search for this same mutation in American and Brazilian women could not detect a single case in women presenting with premature ovarian failure.

As more women with premature ovarian failure undergo genetic studies, we anticipate the identification of multiple subgroups, each with a different biologic basis for the ovarian failure, results that would be consistent with a heterogeneous condition with multiple causes. For example, a case of hypergonadotropic primary amenorrhea has been reported due to a point mutation in the LH receptor gene; FSH and LH were only slightly elevated and multiple ovarian follicles were present with development and steroidogenesis up to an early antral stage. The almost normal gonadotropin levels were probably due to inhibition by inhibin, because inhibin secretion by granulosa cells is FSH-dependent and not influenced by LH. This patient had two siblings who were 46,XY male pseudohermaphrodites due to Leydig cell hypoplasia because of the same LH receptor gene mutation. In another example, translocations between regions on X and Y chromosomes that share sequence homology have been reported in patients with secondary amenorrhea and ovarian failure. Sequences on the long arm of the X chromosome (Xq27-28) share homology with sequences on the long arm of the Y chromosome (Yq11.22) allowing errors in the process of crossing-over. In addition, less than complete mutations (so-called premutations) of the site that transmits the fragile X syndrome have been reported to occur at a greater frequency in women with premature ovarian failure.

The Effect of Radiation and Chemotherapy

The effect of radiation is dependent upon age and the x-ray dose. Steroid levels begin to fall and gonadotropins rise within 2 weeks after irradiation to the ovaries. The higher number of oocytes in younger age is responsible for the resistance to total castration in young women exposed to intense radiation. Function can resume after many years of amenorrhea. On the other hand, the damage may not appear until later in the form of premature ovarian failure. If pregnancy occurs, the risk of congenital abnormalities is no greater than normal. When the irradiation field excludes the pelvis, there is no risk of premature ovarian failure. For this reason, elective transposition by laparoscopy of the ovaries out of the pelvis prior to irradiation provides a good prospect for future fertility. Gonads are not in danger in the kitchen; microwave ovens utilize wavelengths with low tissue-penetrating power.

The following table indicates the risk of sterilization according to dose.

<table>
<thead>
<tr>
<th>Ovarian Dose Sterilization Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 rads No effect</td>
</tr>
<tr>
<td>150 rads Some risk over age 40</td>
</tr>
<tr>
<td>250–500 rads Ages 15–40: 60% sterilized</td>
</tr>
<tr>
<td>500–800 rads Ages 15–40: 60–70% sterilized</td>
</tr>
<tr>
<td>over 800 rads 100% permanently sterilized</td>
</tr>
</tbody>
</table>

Alkylating agents are very toxic to the gonads. As with radiation, there is an inverse relationship between the dose required for ovarian failure and age at the start of therapy. Other chemotherapeutic agents have the potential for ovarian damage, but they have been less well studied. The effect of combination chemotherapies is similar to those of the alkylating agents. Approximately two-thirds of premenopausal women with breast cancer and treated with cyclophosphamide, methotrexate, and fluorouracil lose ovarian function. Resumption of menses and pregnancy can occur, but there is no way to predict which patient will reacquire ovulatory function. As with radiotherapy, damage may present late with premature ovarian failure.

Is it possible that maintenance of ovarian follicles in a "dormant" state by suppressing FSH secretion can prevent ovarian failure? In a monkey model, gonadotroph suppression by treatment with a GnRH agonist during radiation did not protect against loss of ovarian follicles. In contrast, GnRH agonist treatment of monkeys did...
The harvesting and cryopreservation of oocytes prior to irradiation and/or chemotherapy will, in our opinion, ultimately prove to be best means of preserving fertility for these patients.

Compartment III: Disorders of the Anterior Pituitary

A consideration of the disorders of the hypothalamic-pituitary axis must first focus on the problem of the pituitary tumor. Fortunately, malignant tumors are almost never encountered. Through 1997, there were approximately 40 reported cases of prolactin cell metastatic carcinomas (mostly in men) and through 1989, no more than 40 cases of primary pituitary cancer in the world literature. However, because a high blood level (350–400 ng/mL) of prolactin composed predominantly of high molecular weight prolactin has been reported in a woman with amenorrhea and/or galactorrhea, which occurs before the eventful clinical expression of a tumor that secretes ACTH or growth hormone. If clinical criteria suggest Cushing's disease, ACTH levels and the 24-hour urinary levels of free cortisol should be measured, and the rapid suppression test (Chapter 13) should be utilized. If acromegaly is suspected, growth hormone should be measured during an oral glucose tolerance test (lack of suppression of growth hormone levels is diagnostic), and the circulating level of IGF-I should be measured. However, the 2 most common tumors are prolactin-secreting adenomas and clinically nonfunctioning tumors. Though usually a problem in adult life, prolactin-secreting tumors can be seen in preadolescent and adolescent children, and thus, can be a cause of failure of growth and development or of primary amenorrhea.

Because acromegaly can initially present with an elevated prolactin level and amenorrhea, the circulating level of IGF-I should be measured in all patients with a macroadenoma (>10 mm diameter).

The majority of clinically nonfunctioning (null) adenomas (30–40% of all pituitary tumors) are of gonadotroph origin and actively secrete FSH, free a-subunit, and rarely, LH (of which all do not exert clinical effects, and, therefore, patients with these tumors usually present with symptoms due to mass effects). The a-subunit can be used as a tumor marker; however, in postmenopausal women (the age at which most gonadotroph adenomas present) the situation can be confusing because increased free a-subunit secretion accompanies increased secretion of gonadotropins. One premenopausal woman with a gonadotroph adenoma has been reported, distinguished by multiple ovarian cysts, hyperestradiol levels, and endometrial hyperplasia. In contrast to patients with normal pituitary glands, patients with pituitary adenomas usually fail to down-regulate gonadotropin secretion in response to GnRH agonist treatment, and repeated GnRH agonist administration is associated with persistent elevations in either FSH or a-subunit. Most patients with these tumors, however, have reduced secretion of gonadotropins (and amenorrhea) because of tumor compression of the pituitary stalk and interference with the delivery of hypothalamic GnRH. For this reason, these patients often present with modest elevations of prolactin (due to the inability of dopamine to reach the anterior pituitary).

Elevated gonadotropins in the presence of a pituitary microadenoma in a woman with amenorrhea are not a consequence of secretion by the tumor; another explanation must be pursued. Not all intrasellar masses are neoplastic. Cysts, tuberculosis, sarcoidosis, and fat deposits have been reported as causes of pituitary compression leading to hypogonadotropic amenorrhea. Lymphocytic hypophysitis is a rare autoimmune infiltration of the pituitary that can mimic a pituitary tumor, often occurring during pregnancy or in the first 6 months postpartum. In the initial phase of hypophysitis, hyperprolactinemia is common, followed by hypothalamicopaenia. Transsphenoidal surgery is both diagnostic and therapeutic for this potentially lethal condition.

Nearby lesions, such as internal carotid artery aneurysms and obstruction of the aqueduct of Sylvius, can also cause amenorrhea. Pituitary insufficiency can be secondary to ischemia and infarction and appear as a late sequela to obstetrical hemorrhage — the well known Sheehan's syndrome. These problems, as well as genetic disorders such as Laurence-Moon-Biedl and Prader-Willi syndromes, are so rarely encountered that consultation with textbooks and colleagues is necessary. In a large transsphenoidal surgical series, 91% of sellar and parasellar masses were pituitary adenomas. Diabetes insipidus is not associated with pituitary adenomas, but commonly accompanies masses not of pituitary origin.

Treatment of Nonfunctioning Adenomas

If imaging discovers a microadenoma (less than 10 mm in diameter) in an asymptomatic patient, no treatment is necessary. These tumors are often incidental findings. A follow-up imaging is recommended in a year or two to be sure there is no growth. If a macroadenoma (greater than 10 mm in diameter) is present and symptomatic, surgery is necessary; these tumors are commonly not detected until the onset of symptoms (headaches and visual disturbances). Because of their large size and the high risk of recurrence, adjunctive irradiation is recommended if residual tumor and elevated gonadotropins and a-subunit are present after surgery. Follow-up imaging is obtained every 6 months for 1 year, and then yearly for 3–5 years. The radiation dose is high (4500 rads), and the incidence of hypopituitarism may reach 50% over a 10-year period. Ongoing surveillance of adrenal and thyroid function is necessary. With careful imaging follow-up after surgery, postoperative irradiation can be avoided in some patients. Although a response to dopamine agonist therapy has been reported, in general, results are not satisfactory. The response to the somatostatin analogue, octreotide, has also been disappointing. Nevertheless, because good tumor size reduction occasionally occurs, the option of medical treatment should be considered.

Pituitary Prolactin-Secreting Adenomas

Prolactin-secreting adenomas are the most common pituitary tumors, and they account for 50% of all pituitary adenomas identified at autopsy. Classically, pituitary adenomas have been grouped according to their staining ability as eosinophilic, basophilic, or chromophobic. This classification is misleading and of no clinical usefulness. Pituitary adenomas should be classified according to their function; e.g., prolactin-secreting adenoma.

With the utilization of the serum prolactin assay and the increased sensitivity of the new imaging techniques, the association of amenorrhea and small pituitary tumors has become recognized as a relatively common problem. This is not a new phenomenon, rather it reflects more sensitive diagnostic techniques. Attempts to link the problem to oral contraceptive use have proved negative. The exact incidence of this clinical problem is unknown. In autopsy series the percentage of pituitary glands found to contain microadenomas ranged from 9% to 27%.

The age distribution ranged from 2 to 86, with the greatest incidence in the 6th decade of life. The sex distribution was equal. However, clinical manifestations, mainly a disruption of the reproductive mechanism, occur more commonly in women and are probably due to estrogen-induced activity of the pituitary lactotrophs. A high prolactin level is encountered in about one-third of women with no obvious cause of amenorrhea. Only one-third of women with high prolactin levels will have galactorrhea, probably because the low estrogen environment associated with the amenorrhea prevents a normal response to prolactin. Another possible explanation again focuses on the heterogeneity of peptide hormones. Prolactin circulates in various forms with structural modifications, which are the result of glycosylation, phosphorylation, deletions, and additions. The various forms are associated with varying bioactivity (manifested by galactorrhea) and immunoreactivity (recognition by immunoblotting). One predominant variant is little prolactin (80–85%), which also has more biological activity than the larger sized variants. Therefore, it is not surprising that big prolactins compose the major form of circulating prolactin in women with normal menses and minimal galactorrhea. This is not always the case, however, because a high blood level (350–400 ng/mL) of prolactin composed predominantly of high molecular weight prolactin has been reported in a woman with oligomenorrhea and galactorrhea but with no evidence of a pituitary tumor. These high levels of relatively inactive prolactin in the absence of a tumor may be due to the creation of macromolecules of prolactin by antiprolactin autoantibodies. Explanations for clinically illogical situations can be found in the variable molecular heterogeneity of the peptide hormones. At any one point in time, the bioactivity and the immunoreactivity of prolactin represent the cumulative effect of the circulating
family of structural variants.

Very high prolactin levels (greater than 1000 ng/mL) are associated with invasive tumors. These very rare tumors do not yield themselves to surgery, but, fortunately, they can usually be effectively treated and controlled with a dopamine agonist.

Approximately one-third of women with galactorrhea have normal menses. As the prolactin concentration increases, a woman can progress sequentially from normal ovulation to an inadequate luteal phase to intermittent anovulation to total anovulation to complete suppression and amenorrhea.

Probably as many as one-third of patients with secondary amenorrhea will have a pituitary adenoma, and if galactorrhea is also present, half will have an abnormal sella turcica. The clinical symptoms do not always correlate with the prolactin level, and patients with normal prolactin levels can have pituitary tumors. The highest prolactin levels, however, are associated with amenorrhea, with or without galactorrhea.

The amenorrhea associated with elevated prolactin levels is due to prolactin inhibition of the pulsatile secretion of GnRH. The pituitary glands in these patients respond normally to GnRH, or in augmented fashion (perhaps due to increased stores of gonadotropins), thus indicating that the mechanism of the amenorrhea is a decrease in GnRH. Short-term administration of an opioid antagonist suggests that this inhibition is mediated by increased opioid activity. However, chronic administration of naltrexone (a long-acting opioid antagonist) does not restore menstrual function. Nevertheless, treatment that lowers the circulating levels of prolactin restores ovarian responsiveness and menstrual function. This is true whether the treatment consists of removal of a prolactin-secreting tumor or suppression of prolactin secretion. Interestingly, postmenopausal women with elevated levels of prolactin do not experience vasomotor symptoms (hot flushes) until prolactin levels are restored to normal.

The increased ability to detect pituitary tumors has been accompanied by the development of a surgical technique that effectively removes the small tumors with a high margin of safety. Utilizing the operating microscope, the transsphenoidal technique approaches via a sublabial incision (under the upper lip), with dissection under the nasal mucosa, removal of the nasal septum to expose the sphenoidal sinus, and resection of the floor of the sphenoid sinus to expose the sella turcica. Tumor tissue is usually distinguishable from the surrounding normal tissue and is often vague. The ideal time for excision is when the adenoma is a small nodule. When enlarged, it becomes more difficult to distinguish normal from pathological tissue. Once the adenoma grows beyond the sela, total removal is essentially impossible.

The development of transsphenoidal surgery was paralleled by the availability and clinical application of the drug, bromocriptine, which specifically suppresses prolactin secretion. Initially, appropriate decisions between the two approaches were difficult to make. With increasing experience, clinical perspective has been achieved, and reasonable judgments are now possible. Let us first consider results with surgery, and then examine dopamine agonist treatment.

Results With Surgery

Transsphenoidal neurosurgery achieves immediate resolution of hyperprolactinemia with resumption of cyclic menses in approximately 30% of patients with macroadenomas and 70% of patients with microadenomas. Besides an inability to achieve a complete cure, surgery may be followed by recurrence of tumor (long-term cure rate is about 55%), ranging from as high as 75% for microadenomas to as low as 10% for macroadenomas), depending upon the skill and experience of the neurosurgeon and the size of the tumor, and a still unknown but significant percentage (perhaps as high as 10–30% after surgery for macroadenomas) of development of panhypopituitarism. Other complications of surgery include cerebrospinal fluid leaks, an occasional case of meningitis, and the frequent postoperative problem of diabetes insipidus. The diabetes insipidus is usually a transient problem, rarely lasting as long as 6 months, but it can be permanent. There is a mortality rate of less than 1%. Although initial follow-up reports of the results of transsphenoidal adenomectomy were discouraging (high recurrence rates), other authors have argued that surgical techniques improved with time, and recurrent hyperprolactinemia is relatively low. The best results are in patients with prolactin levels in the 150–500 nmol/L range; the higher the prolactin the lower the cure rate. In the largest series (409 women) with long-term follow-up, the recurrence rate was 26% when postoperative prolactin levels were 20 nmol/L or less. The higher the postoperative prolactin level, the lower the cure rate. Overall, approximately 50% of both microadenomas and macroadenomas cured by surgery. There were no deaths, but adrenal hormone replacement (hypopituitarism) was necessary in 4%. Importantly, pregnancy (achieved by 88% within 1 year in those desiring conception) did not cause exacerbation or recurrence in a single patient.

There are 3 possible explanations for the recurrence or persistence of hyperprolactinemia after surgery.

1. The prolactin-producing tumor looks like the surrounding normal pituitary, and it is difficult to resect completely.
2. The tumor may be multifocal in origin.
3. There may be a continuing abnormality of the hypothalamus giving rise to chronic stimulation of the lactotrophs. In other words, this is a problem of recurrent hyperprolactinemia.

When persistent or recurrent tumors are treated with repeat surgery, only about one-third are cured and the complication rate is high. Irradiation after surgery has many problems, including the possibility of stroke and other brain tumors. Irradiation after surgery should be reserved as adjunctive therapy for controlling postoperative persistence or regrowth of large tumors and shrinking large tumors that are unresponsive to medical treatment. Overall, a small number of irradiated women return to normal hormonal function. 

Results With Radiation

Results with radiation therapy are less satisfactory than with surgery. In addition, response is very slow; prolactin concentrations may take several years to fall. After radiation, panhypopituitarism can occur as long as 10 years after treatment. Patients who have been treated with radiation should be followed for a long time, and any symptoms suggestive of pituitary failure require investigation. It is not yet known whether focusing the delivery of radiation (stereotactic radiotherapy) can avoid the overall, approximately 50% of both microadenomas and macroadenomas are cured by surgery. Furthermore, it is believed that focused irradiation is better for small tumors or small amounts of residual tumor after surgery. Irradiation should be reserved as adjunctive therapy for controlling postoperative persistence or regrowth of large tumors and shrinking large tumors that are unresponsive to medical treatment. Overall, a small number of irradiated women return to normal hormonal function.

Dopamine Agonist Treatment

Bromocriptine is a lysergic acid derivative with a bromine substitute at position 2. It is available as the methane-sulfonate (mesylate) in 2.5 mg tablets. It is a dopamine agonist, binding to dopamine receptors and, therefore, directly mimicking dopamine inhibition of pituitary prolactin secretion. Prolactin-secreting tumors do not have inactivating mutations in the dopamine receptor gene; thus, dopamine agonists can bind and exert inhibiting activity. Absorption from the gastrointestinal tract is rapid but not complete; 28% is absorbed and 94% metabolized in the first pass through the liver. Bromocriptine is metabolized into at least 30 excretory products. Excretion is mainly biliary, and more than 90% appears in the feces over 5 days after a single dose of 2.5 mg. A small part, 6–7%, is excreted unchanged or as metabolites in the urine.

The oral dose that suppresses prolactin is 10 times lower than that which improves the symptoms of Parkinson's disease. For some patients, one pill a day (or half a pill bid) will be effective. On the other hand, an occasional patient will require 7.5 mg or 10 mg daily in order to suppress adenaoma secretion of prolactin.

Bromocriptine is also available in a long-acting form (depot-bromocriptine) for intramuscular injection and as a slow-release oral form. Depot-bromocriptine is administered in a dose of 50–75 mg/monthly; the dose of the oral slow-release formulation is 5–15 mg daily. These forms are equally effective as the standard oral...
Macroadenomas
Summary: Therapy of Pituitary Prolactin-Secreting Adenomas

The response of macroadenomas to bromocriptine is impressive, and a most compelling reason in favor of its use is that it has been successful when previous surgery or radiation has failed. The problem, however, is that it probably must be taken indefinitely, because there is yet to be a convincing report of complete disappearance and resolution of the tumor that can be attributed to drug therapy and not spontaneous resolution. Light and electron microscopic, immunohistochemical, and neurochemical analyses all indicate that bromocriptine causes not only a reduction in the size of individual cells but also necrosis of the cells with replacement fibrosis. Nevertheless, prolactin levels generally return to an elevated state after discontinuation of the drug. There are cases of improvement in sellar imaging; however, the occurrence of spontaneous regression of prolactin-secreting tumors makes it impossible to attribute "cures" to bromocriptine. Recurrence of hyperprolactinemia has been observed following as many as 4–8 years of treatment.

Other Dopamine Agonists
Other ergoline derivatives with dopaminergic activity are available throughout the world. Pergolide is more potent, longer-lasting, and better tolerated by some patients than bromocriptine. Pergolide is given in a single daily dose of 50–150 mg, and it may be effective in bromocriptine-resistant patients. Others are lysuride, terguride, metergoline, and cabergoline. Quinagolide (CV 205–502) is a nonergot long-acting dopamine agonist given in a daily bedtime dose of 75–300 mg. Because quinagolide has a higher affinity for the dopamine receptor, tumors resistant to bromocriptine have responded to this drug. Side effects are reduced with quinagolide, and it appears to have antidepressive properties. Side effects and intolerance with one of these drugs is often solved by using another.

Cabergoline. Cabergoline is also an ergot-derivative dopamine agonist. Patients resistant to both bromocriptine and quinagolide have been reported to respond to cabergoline. Cabergoline can be administered orally at doses of 0.5 to 3.0 mg only once weekly, although it can be given twice weekly if necessary, with minimal dopamine agonist side effects (headache being the most common complaint). The low rate of side effects and the once weekly dosage make cabergoline an attractive choice for initial treatment, replacing bromocriptine. The only reservation (a small one) is a more limited experience documenting fetal safety for patients being treated for infertility. Cabergoline can also be administered vaginally for the rare patient who cannot tolerate it orally.

Summary: Therapy of Pituitary Prolactin-Secreting Adenomas
Macrodenomas

Approximately 10% of patients cannot tolerate oral bromocriptine. Nausea, headache, and faintness are the usual initial problems. The faintness is due to orthostatic hypotension, which can be attributed to relaxation of smooth muscle in the splanchnic and renal beds, as well as inhibition of transmitter release at noradrenergic nerve endings and central inhibition of sympathetic activity. Neuropsychiatric symptoms, occasionally with hallucinations, occur in less than 1% of patients. This may be due to hydrolysis of the lysergic acid part of the molecule. Other side effects include dizziness, fatigue, nasal congestion, vomiting, and abdominal cramps.

Side effects can be minimized by slowly building tolerance toward the usual dose. 2.5 mg bid. Treatment should be started with an initial dose of 2.5 mg given at bedtime. It may help to take the tablet with a glass of milk and snack. The peak level is achieved 2 hours after ingestion, and the biological half-life is about 3 hours. If intolerance occurs with this initial dose, then the tablet should be cut in half, and an even slower program should be followed. Usually a week after the initial dose, the second 2.5 mg dose can be added at breakfast or lunch. Patients who are extremely sensitive to the drug should be instructed to divide the tablets and to devise their own schedule of increasing dosage in order to achieve tolerance. A very small percentage of patients cannot tolerate any dosage.

Vaginal administration of bromocriptine is an excellent method to avoid side effects. One 2.5 mg tablet is inserted high into the vagina at bedtime. This dose will provide excellent clinical results and few side effects. In contrast to oral bromocriptine which is not absorbed completely and that which is absorbed is largely metabolized in the first pass through the liver, vaginal absorption is nearly complete, and avoidance of the liver first-pass effect (with longer maintenance of systemic levels) allows achievement of therapeutic results at a lower dose.

There are two bromocriptine treatment methods to follow in those patients seeking pregnancy. The first is simply daily administration of 2.5 mg bid until the patient is pregnant as judged by the basal body temperature chart. In the second method, bromocriptine is administered during the follicular phase, and the drug is stopped when a basal body temperature rise indicates that ovulation has occurred, thus avoiding high drug levels early in pregnancy. The drug is resumed at menses when it is apparent the patient is not pregnant. No comparative study has been performed to tell us whether the follicular phase only method is as effective as the daily method. Furthermore, there has been no evidence that bromocriptine ingestion during early pregnancy is harmful to the fetus.

Results of Treatment. In 22 clinical trials, 80% of patients with amenorrhea/galactorrhea, associated with hyperprolactinemia but no demonstrable tumors, had menses restored. The average treatment time to the initiation of menses was 5.7 weeks. Complete cessation of galactorrhea occurred in 50–60% of patients in an average time of 12.7 weeks, and a 75% reduction of breast secretions was achieved in 6.4 weeks. It is important to advise patients that the cessation of galactorrhea is a slower and less certain response than restoration of ovulation and menses. Amenorrhea recurred in 41% of the patients within an average of 4.4 weeks of discontinuing treatment; galactorrhea recurred in 69% at an average of 6.0 weeks. Approximately 5% of patients terminated treatment because of adverse reactions.

Regression of Tumors With Bromocriptine. There is no question that macroadenomas will regress with bromocriptine treatment. In some, there is prompt shrinkage with low-dose treatment (5–7.5 mg daily); in others, prolonged treatment is required with higher doses. If a prolactin adenoma fails to shrink with 10 mg daily, further increases in dose are not useful. Visual improvement may be noted within several days. Reduction in tumor size can take place in several days to 6 weeks, but in some cases it is not observed until 6 months or more. In most cases, rapid shrinkage occurs during the first 3 months of therapy, followed by slower reduction. Very high prolactin levels, greater than 2000–3000 ng/mL, are probably the result of invasion of the cavernous sinuses with release directly into the bloodstream. Levels greater than 1000 ng/mL are associated with locally invasive tumors. Even these cases show remarkable resolution with bromocriptine treatment. Indeed, surgical results with invasive tumors are so poor that long-term control with a dopamine agonist is recommended. Although tumor shrinkage is always preceded by a decrease in prolactin levels, the overall response cannot be predicted by the basal prolactin level, the absolute or relative fall in prolactin, or even the attainment of normal prolactin levels. Visual improvement improves rapidly, but maximal effects may take several months. However, a prolactin level nonresponder will be a tumor size nonresponder.

The response of macroadenomas to bromocriptine is impressive, and a most compelling reason in favor of its use is that it has been successful when previous surgery or radiation has failed. The problem, however, is that it probably must be taken indefinitely, because there is yet to be a convincing report of complete disappearance and resolution of the tumor that can be attributed to drug therapy and not spontaneous resolution. Light and electron microscopic, immunohistochemical, and neurochemical analyses all indicate that bromocriptine causes not only a reduction in the size of individual cells but also necrosis of the cells with replacement fibrosis. Nevertheless, prolactin levels generally return to an elevated state after discontinuation of the drug. There are cases of improvement in sellar imaging; however, the occurrence of spontaneous regression of prolactin-secreting tumors makes it impossible to attribute "cures" to bromocriptine. Recurrence of hyperprolactinemia has been observed following as many as 4–8 years of treatment.
Currently dopamine agonist treatment is advocated for the treatment of macroadenomas, utilizing as low a dose as possible. Shrinkage of a tumor may require 5–10 mg bromocriptine daily, but once shrinkage has occurred, the daily dose should be progressively reduced until the lowest maintenance dose is achieved. The serum prolactin level can be utilized as a marker, checking levels every 3 months until stable. In many (but not all) patients, control of tumor growth correlates with maintenance of a baseline prolactin level and can be achieved in some patients with as little as one-quarter of a tablet (0.625 mg) daily. Withdrawal of the drug is usually associated with regrowth or reexpansion of the tumor, and, therefore, treatment must be long-term if not indefinite. If there is a good response in prolactin levels, and if present, visual field defects, the MRI should be repeated after 1 year of treatment to establish size reduction of the tumor. Some patients will prefer surgery rather than long-term medical treatment, and it is certainly a legitimate option. In view of better results claimed in more recent times, this choice should be presented to the patient. Transsphenoidal surgery is recommended when suprasellar extension or visual impairment persists after dopamine agonist treatment of a macroadenoma. For some patients, side effects with dopamine agonist treatment and difficulty with medication compliance make surgery a reasonable alternative. Because tumor recurrence is high, radiotherapy should be considered. All patients receiving radiotherapy require ongoing surveillance for the development of hypopituitarism. Surgery should be considered as a debulking procedure for very large tumors with or without invasion prior to long-term dopamine agonist therapy.

Short-term treatment (several weeks) with a dopamine agonist can make surgery easier because of a reduction in size (although not all neurosurgeons agree); however, long-term treatment (3 months or more) is associated with fibrosis, which makes complete surgical removal more difficult and more likely to be associated with the sacrifice of other pituitary hormonal function. Even though prolactin levels usually increase when dopamine agonist treatment is discontinued after several years, many tumors (70–80%) do not regrow. Pregnancy should be deferred until repeat imaging confirms shrinkage of the macroadenoma.

Approximately 10% of macroadenomas do not shrink with dopamine agonist therapy. The failure of a tumor to shrink significantly in size despite a normalization of prolactin levels is consistent with a nonfunctioning tumor that is interrupting the supply of dopamine to the pituitary by stalk compression. Early surgery is indicated. A tumor that continues to grow despite dopamine agonist treatment may be a rare carcinoma.

Microadenomas

The treatment of microadenomas should be directed to alleviating one of two problems: infertility or breast discomfort. Treatment with a dopamine agonist is the method of choice. Again, some patients, deliberately and understandably, choose the surgical approach in hopes of achieving a cure and avoiding the worry and annoyance of continuing surveillance.

The major therapeutic dilemma can be expressed by the following question: should chronic dopamine agonist treatment be utilized to retrieve ovarian function in those patients with hypoestrogenic amenorrhea, or should estrogen treatment be offered? Until a clear-cut benefit is demonstrated by clinical studies, we cannot advocate widespread dopamine agonist therapy for those patients not interested in becoming pregnant. This conservative approach is supported by documentation of a benign clinical course with spontaneous resolution in many patients. Patients with hypoestrogenic amenorrhea are encouraged to be on an estrogen therapy program to maintain the health of their bones and the vascular system. Low-dose oral contraception is recommended for those patients who require contraception. Estrogen-induced tumor expansion or growth has not been a problem in both our experience and in that of others.

Long-Term Follow-Up

Because these tumors can grow slowly, it is appropriate in the absence of symptoms to evaluate patients with microadenomas annually for 2 years. The evaluation consists of a measurement of the prolactin level and a coned-down view of the sella turcica. If the course is unchanged, annual evaluation can be limited to measurement of the prolactin level. More sophisticated sellar imaging is reserved for patients with a change in the coned-down x-ray view, an increasing prolactin, or the development of headaches and/or visual complaints. It should be noted that progressively increasing prolactin levels have been observed without associated tumor growth of a microadenoma. The rare microadenoma that grows deserves treatment. Patients with macroadenomas deserve an initial period of follow-up after treatment every 6 months, and if the adenoma appears to be clinically stable, prolactin levels should be measured annually. CT scanning or MRI is reserved for situations suggestive of tumor expansion. If clinician and patient need reassurance regarding tumor size, imaging intervals can be prolonged if the tumor is stable; e.g., at 1 year, 2 years, 4 years, 8 years. Tumor expansion and recurrent tumors after surgery or radiotherapy deserve a trial of treatment with a dopamine agonist.

Patients who have been on dopamine agonist treatment for 5–10 years with successful tumor size reduction can have a gradual reduction, and then stopping of treatment, followed by monitoring of prolactin levels every 3 months. If a normal prolactin level is maintained, we recommend an imaging study 1 year later. Of course, tumor reexpansion requires resumption of treatment with the gradual program that should always be used when starting therapy.

Pregnancy and Prolactin Adenomas

Approximately 80% of hyperprolactinemic women achieve pregnancy with dopamine agonist treatment. Breastfeeding, if desired, can be experienced normally without fear of stimulating tumor growth. There is no increase in the normal rates of spontaneous miscarriages, ectopic pregnancies, and complications of pregnancy.

Interestingly, some women resume cyclic menses after pregnancy. This spontaneous improvement may be due to tumor infarction brought about by the expansion and shrinkage during and after pregnancy, or there may be a correction of a hypothalamic dysfunction followed by a disappearance of the associated pituitary hyperplasia.

A very small percentage (less than 2%) of women with hyperprolactinemia and microadenomas will develop signs or symptoms suggestive of tumor growth during pregnancy. Approximately 5% of these patients will develop asymptomatic tumor enlargement (determined by radiologic techniques), and essentially none will ever require surgical intervention. The risk is higher with macroadenomas, approximately 15%. Heads usually precede visual disturbances, and both may occur in any trimester. There is no characteristic headache; they are variable in intensity, location, and character. Bitemporal hemianopsia is the classic visual field finding, but other defects can occur. It has been argued in the past that a desire for pregnancy was a reason for the surgical approach. This argument hinged on the risk of tumor enlargement during pregnancy due to the well known stimulatory effects of estrogen on the pituitary lactotrophs. As noted above, however, experience has indicated that very few patients develop problems. In a series of 65 consecutive pregnant women with untreated pituitary adenomas, not one with a macroadenoma developed visual loss; however, 6 of 8 women with macroadenomas did develop some visual field loss.

It is impossible to identify which patient is at risk for symptomatic expansion during pregnancy. Other than a very large tumor, the size is not critical because both microadenomas and macroadenomas can undergo uneventful pregnancies. There is no increase in miscarriages, or perinatal mortality or morbidity. It is virtually
unheard of to develop a problem that results in perinatal damage or serious maternal sequelae. Nevertheless, an occasional serious event can occur; e.g., hemorrhage of the tumor with diabetes insipidus and the potential for permanent visual impairment or life-threatening sequelae. 122 Surveillance during pregnancy at first consisted of monthly visual field and prolactin measurements. With experience, this proved to be unnecessary. The patient and the clinic can be guided by the development of symptoms. Assessment of visual fields, prolactin, and the sella turcica by imaging can await the onset of headaches or visual disturbances. Even macroprolactinomas with suprasellar extension can be followed closely; discontinuation of a dopamine agonist after conception is usually not associated with tumor growth during the pregnancy. 123 With repeated pregnancies, tumor regression often occurs with the development of an empty sella.

Definite evidence of tumor expansion, as well as the symptoms of headaches and visual changes, promptly regresses with dopamine agonist treatment. 120 Termination of pregnancy or neurosurgery, therefore, should rarely, if ever, be necessary. Although bromocriptine treatment profoundly lowers both maternal and fetal blood levels of prolactin, no adverse effects on the pregnancy or the newborn have been noted. 123, 175, 180, 181 and 122 Fortunately, amniotic fluid prolactin (and its presumed action on regulation of amniotic fluid water and electrolytes) is derived from decidual tissue, and its secretion is controlled by estrogen and progesterone, not dopamine. Therefore, dopamine agonist treatment does not affect amniotic fluid levels of prolactin.

The Empty Sella Syndrome

A patient may have an abnormal sella turcica, but rather than a tumor, she may have the empty sella syndrome. In this condition, there is a congenital incompleteness of the sellar diaphragm that allows an extension of the subarachnoid space into the pituitary fossa. The pituitary gland is separated from the hypothalamus and is flattened. The sella floor may be demineralized due to pressure from the cerebrospinal fluid, and the x-ray picture on coned-down views will be similar to a tumor. The empty sella syndrome can also occur secondary to surgery, radiotherapy, or infarction of a pituitary tumor.

An empty sella is found in approximately 5% of autopsies, and approximately 85% are in women, previously thought to be concentrated in middle-aged and obese women. 121 A closer look at the sella turcica, brought about by our pursuit of elevated prolactin levels, has revealed an incidence of empty sellas in 4–16% of patients who present with amenorrhea/galactorrhea. 122, 126 Galactorrhea and elevated prolactin levels can be seen with an empty sella, and there may be a coexisting prolactin-secreting adenoma. This suggests that the empty sella in these patients may have arisen because of tumor infarction.

This condition is benign; it does not progress to pituitary failure. The chief hazard to the patient is inadvertent treatment for a pituitary tumor. Even though enlargement of the sella turcica with a normal shape is more likely associated with an empty sella than a tumor, all patients should have examination by imaging for confirmation.

Because of the possibility of a coexisting adenoma, patients with elevated prolactin levels or galactorrhea and an empty sella should undergo annual surveillance (prolactin assay and coned-down view) for a few years to detect tumor growth. It is totally safe and appropriate to offer hormone treatment or induction of ovulation.

Sheehan's Syndrome

Acute infarction and necrosis of the pituitary gland due to postpartum hemorrhage and shock is known as Sheehan's syndrome. 124 The symptoms of hypopituitarism are usually seen early in the postpartum period, especially failure of lactation and loss of pubic and axillary hair. Deficiencies in growth hormone and gonadotropins are most common, followed by ACTH, and last, by TSH in frequency. 125 Diabetes insipidus is not usually present. This can be a life-threatening condition, but fortunately, because of good obstetrical care, this syndrome is never encountered by most of us.

Compartment IV: Central Nervous System Disorders

Hypothalamic Amenorrhea

Patients with hypothalamic amenorrhea (hypogonadotrophic hypogonadism) have a deficiency in GnRH pulsatile secretion. Hypothalamic problems are usually diagnosed by exclusion of pituitary lesions and are the most common category of hypogonadotrophic amenorrhea, a functional suppression of reproduction, often a psychobiologic response to life events. 149 Frequently, there is an association with a stressful situation, such as in business or in school. There is also a higher proportion of underweight women and a higher occurrence of previous menstrual irregularity. Indeed, many women with hypothalamic amenorrhea display the endocrine and metabolic characteristics associated with eating disorders and athletics, suggesting the presence of a subclinical eating disorder. 127 Nevertheless, the clinician is obliged to go through the process of exclusion prior to prescribing hormone therapy or attempting induction of ovulation to achieve pregnancy.

The degree of GnRH suppression determines how these patients present clinically. Mild suppression can be associated with a marginal effect on reproduction, specifically an inadequate luteal phase. Moderate suppression of GnRH secretion can yield anovulation with menstrual irregularity, and profound suppression is manifested by hypothalamic amenorrhea.

Patients with hypothalamic amenorrhea are categorized by low or normal gonadotropins, normal prolactin levels, a normal imaging evaluation of the sella turcica, and a failure to demonstrate withdrawal bleeding. A good practice is to evaluate such patients annually. This annual surveillance should include a prolactin assay and the coned-down view of the sella turcica. The x-ray is necessary only every 2–3 years after several years with no change. In the only long-term follow-up of a large group of women with secondary amenorrhea, it was noted that amenorrhea associated with psychological stress or weight loss demonstrated a spontaneous recovery after 6 years in 72% of the women. 125 This still leaves a significant percentage of women who require ongoing surveillance. In patients with eating disorders, the return of menstrual function is associated with a gain in body weight, an obvious marker for clinical improvement in the underlying condition. 126 Experimental evidence in the monkey indicates that corticotropin-releasing hormone (CRH) inhibits gonadotropin secretion, probably by augmenting endogenous opioid secretion. 144 This is the probable pathway by which stress interrupts reproductive function. Women with hypothalamic amenorrhea have reduced secretion of FSH, LH, and prolactin, but increased secretion of cortisol. 125, 149 and 122 There is also evidence to indicate that some patients with hypothalamic amenorrhea have dopamineergic inhibition of GnRH pulse frequency. 144 The suppression of GnRH pulsatile secretion may be the result of increases in both endogenous opioids and dopamine. Thus far, abnormalities in the genes for GnRH and the beta-subunits for FSH and LH have not been detected in patients with hypothalamic amenorrhea. 130, 146

Even though a patient may not be currently interested in pursuing pregnancy, it is important to assure these patients that, at the appropriate time, treatment for the induction of ovulation will be available and that fertility can be achieved. Concern with potential fertility is often an unspoken fear, especially in the younger patients, even teenagers. On the other hand, induction of ovulation should be performed only for the purpose of producing a pregnancy. There is no evidence that cyclic hormone administration or induction of ovulation will stimulate the return of normal function.

Weight Loss, Anorexia, Bulimia

St. Wilgefortis was the 7th daughter of the King of Portugal, living around the year 1000. 127 When confronted with an arranged marriage (she had made a vow of virginity to become a nun), she turned to intense prayer. The intensity of the prayer was marked by anorexia and the growth of body hair. Confronted with this new appearance, the King of Sicily changed his mind about the marriage, and Wilgefortis's father had her crucified. Around 1200, the legend of Wilgefortis spread throughout Europe.

St. Wilgefortis became a symbol, a woman who liberated herself from female problems, and she became a protectress of women with sexual problems, including problems associated with childbirth. Indeed, women who wished to rid themselves of their husbands prayed to her, because she had successfully resisted both a father and a potential husband. In England, rather than St. Wilgefortis, she was known as St. Uncumber because women believed she could uncumber them of their husbands.

Thus emerged the medieval dark ages explanation (with ascendancy to sainthood) of a girl's response (anorexia nervosa) to her fears of marriage and sexuality. Our understanding of the reason for this extraordinary behavior continues today to focus on an inability to cope with the onset of adult sexuality, with a return to the
prepubertal state. Both anorexia nervosa and bulimia nervosa (binge eating) are distinguished by a morbid fear of fatness.

Obesity can be associated with amenorrhea, but amenorrhea in an obese patient is usually due to anovulation, and a hypogonadotropic state is not encountered unless the patient also has a severe emotional disorder. Conversely, acute weight loss, in some unknown way, can lead to the hypogonadotropic state. Again, the clinician must pursue the presence of a pituitary tumor, and the diagnosis of hypothalamic amenorrhea is made by exclusion.

Clinically a spectrum is encountered that ranges from a limited period of amenorrhea associated with a crash diet, to the severely ill patient with the life-threatening attrition of anorexia nervosa. It is a common experience for a clinician to be the first to recognize anorexia nervosa in a patient presenting with the complaint of amenorrhea. It is also not infrequent that a clinician will evaluate and manage an infertility problem due to hypogonadotropism and not be aware of a developing case of anorexia. Because the mortality rate associated with this syndrome is significant (5–15%), it warrants close attention.

Diagnosis of Anorexia Nervosa

1. Onset between ages 10 and 30.
2. Weight loss of 25% or weight 15% below normal for age and height.
3. Special attitudes:
   - Denial
   - Distorted body image
   - Unusual hoarding or handling of food.
4. Least one of the following:
   - Amenorrhea.
   - Other characteristics:
     - Hypercarotenemia,
     - Diabetes insipidus.

Anorexia nervosa has been recognized to occur frequently in young white middle to upper class females under age 25, but it is now apparent that this problem occurs at all socioeconomic levels in about 1% of young women. The families of anorectics are success-achievement-appearance oriented. Serious problems may be present within the family, but the parents make every effort to maintain an apparent marital harmony, glossing over or denying conflicts. In one psychiatric interpretation, each parent, in secret dissatisfaction with the other, expects affection from their "perfect" child. Anorexia begins when the role of the perfect child becomes too difficult. The pattern usually starts with a voluntary diet to control weight. This brings a sense of power and accomplishment, soon followed by a fear that weight cannot be controlled if discipline is allowed to falter. A reasonable view is to consider anorexia as a mechanism that identifies a generally disturbed family. The symptom pattern is the expression of the various psychological, familial, and cultural factors involved.

At puberty, the normal weight gain may be interpreted as excessive, and this can trip the teenager over into true anorexia nervosa. Excessive physical activity can be the earliest sign of incipient anorexia nervosa. The children are characteristically overachievers and strivers. They seldom give any trouble, but are judgmental and demand that others live up to their rigid value system, often resulting in social isolation. Patients with eating disorders usually demonstrate delayed psychosexual development marked by sexual experiences occurring at a later age.

The cultural value our society places on thinness definitely plays a role in eating disorders. Both occupational and recreational environments that stress thinness put women at greater risk for anorexia nervosa and bulimia. But basically, an eating disorder is a method being utilized to solve a psychological dilemma.

Besides amenorrhea, constipation, often severe and accompanied by abdominal pain. The preoccupation with food may manifest itself by large intakes of lettuce, raw vegetables, and low-calorie foods. Hypotension, hypothermia, rough dry skin, soft lanugo-type hair on the back and buttocks, bradycardia, and edema are the most commonly encountered signs. Long-term diuretic and laxative abuse may produce significant hypokalemia. An elevation of the serum carotene is not always associated with a large intake of yellow vegetables, suggesting that a defect in vitamin A utilization is present. The yellowish coloration of the skin is seen on the palms. Hypercarotenemia should be regarded as a metabolic marker, but not every woman with hypercarotenemia will be amenorrheic or anovulatory.

Bulimia is a syndrome marked by episodic and secretive binge eating followed by self-induced vomiting, fasting, or the use of laxatives and diuretics. It appears to be a growing problem among young women; however, careful study indicates that although bulimic behaviors may be relatively common, clinically significant bulimia is not (approximately 1.0% of female students and 0.1% of male students in a college sample), and the overall prevalence of eating disorders may be declining. Bulimic behavior is frequently seen in patients with anorexia nervosa (about half), but not in all. Patients with bulimia have a high incidence of depressive symptoms, and a problem with shoplifting (usually food). Little is known about the long-term effects of bulimic behaviors. There is a growing tendency to divide patients with anorexia nervosa into bulimic anorectics and dieters. Bulimic anorectics are older, less isolated socially, and have a higher incidence of family problems. Body weight in a "pure" bulimic fluctuates, but it does not fall to the low levels seen in anorectics.

The serious case of anorexia nervosa is seen more often by an internist. However, the borderline anorectic frequently presents to a gynecologist, pediatrician, or family physician as a teenager who has low body weight, amenorrhea, and hyperactivity (excellent grades and many extracurricular activities). The amenorrhea can precede, follow, or appear coincidentally with the weight loss.

The various problems associated with anorexia nervosa reflect dysfunction of the body mechanisms regulated by the hypothalamus: appetite, thirst and water conservation, temperature, sleep, autonomic balance, and endocrine secretion. Endocrine studies can be summarized as follows: FSH and LH levels are low, cortisol levels are elevated, prolactin levels are normal, TSH and thyroxine (T4) levels are normal, but the 3.5,3'-triodothyronine (T3) level is low, and reverse T3 is high. Indeed, many of the symptoms can be explained by relative hypothyroidism (constipation, cold intolerance, bradycardia, hypotension, dry skin, low metabolic rates, hypercarotenemia). There appears to be a compensation to the state of undernourishment, with diversion from formation of the active T3 to the inactive metabolite, reverse T3. With weight gain, all of the metabolic changes revert to normal. Even though normal gonadotropin secretion may be restored with weight gain, 30% of patients remain amenorrheic, a good sign of ongoing psychological conflict.

The central origin for the amenorrhea is suggested by the demonstration that the response to GnRH is regained at approximately 15% below the ideal weight, and this return to normal responsiveness occurs before the resumption of menses. Patients with anorexia nervosa have persistent low levels of gonadotropins similar to prepubertal children. With weight gain, sleep-associated episodic secretion of LH appears, similar to the early pubertal child. With full recovery, the 24-hour pattern is similar to that of an adult, marked by fluctuating peaks. This sequence of changes with increasing and decreasing weight is explained by increasing and decreasing pulsatile secretion of GnRH. Neuropeptide Y may be a link between the control of food intake and GnRH secretion. Neuropeptide Y cell bodies are located in the arcuate nucleus of the hypothalamus. This peptide both stimulates feeding behavior and inhibits gonadotropin secretion (presumably by suppressing GnRH pulses, although a direct action on the pituitary is also possible). In response to food deprivation, the endogenous levels of neuropeptide Y increase, and elevated concentrations of neuropeptide Y can be measured in the cerebrospinal fluid of anorexic women. This is consistent with the known actions of leptin, as discussed with full references in Chapter 19.

This is one of the rare conditions in which gonadotropins may be undetectable (large pituitary tumors and genetic deficiencies are the others). If necessary, a high plasma cortisol can differentiate this condition from pituitary insufficiency. However, extensive laboratory testing in these patients is not necessary. Adherence to our scheme for the evaluation of amenorrhea is indicated to rule out other pathological processes. Further endocrine assessment, however, is not essential for patient
A careful and gentle revelation to the patient of the relationship between the amenorrhea and the low body weight is often all that is necessary to stimulate the patient to return to normal weight and normal menstrual function. Occasionally, it is necessary to see the patient frequently and become involved in a program of daily calorie counting (a minimum intake of 2600 calories) in order to break the patient's established eating habits. If progress is slow, hormone therapy should be initiated. In an adult weighing less than 100 pounds, continued weight loss requires psychiatric consultation. Some would argue that any patient with an eating disorder requires psychiatric intervention.

Going away to school or the development of a relationship with a male friend often are turning points for young women with mild to moderate anorexia. A failure to respond to these life changes is relatively ominous, predicting a severe problem with a protracted course.

It is disappointing that despite the impressive studies on anorexia, there is no specific or new therapy available. This only serves to emphasize the need for early recognition to allow psychological intervention before the syndrome is entrenched in its full severity. 52 The use of serotonin antagonists or uptake blockers is restricted to patients who, as is frequently the case, have the comorbidity of clinical depression. 52 Clinicians (and parents) should pay particular attention to weight and diet in young women with amenorrhea. Even in amenorrheic adolescents of normal or above-normal body weights, disordered eating patterns (faming and purging) are often present, a sign of an underlying stressful disorder. 54

Exercise and Amenorrhea

Soranus of Ephesus in the 1st 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. Late in the 20th century, there was a new awareness that competitive female athletes, as well as women engaged in strenuous recreational exercise and women engaged in other forms of demanding activity, such as ballet and modern dance, have a significant incidence of menstrual irregularity and amenorrhea, in the pattern called hypothalamic suppression. The extent of this problem has perhaps been underestimated because of a lack of attention to anovulatory cycles. As many as two-thirds of runners who have menstrual periods have short luteal phases or are anovulatory. 55 When training starts before menarche, menarche can be delayed by as much as 3 years, and the subsequent incidence of menstrual irregularity is higher. In some individuals, secondary amenorrhea is associated with delayed menarche even though training did not begin until after menarche. It is suggested that some girls with these characteristics may be socially influenced to pursue athletic training. Contrary to the female situation, exercise has little effect on the timing of puberty in boys. Although changes in testicular function can be demonstrated in males, the changes are more subtle and less meaningfully clinically. 56

There appear to be two major influences: a critical level of body fat and the effect of stress itself. Young women who weigh less than 115 pounds and lose more than 10 pounds while exercising are the women most likely to develop the problem, 57 an association that supports the critical weight concept of Frisch. 58

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. In dealing with patients, it is helpful to use the nomogram derived from Frisch, which is based on the calculation of the amount of total body water as a percentage of body weight. This relates to the percentage of body fat and, therefore, is an index of fatness. The 10th percentile at age 16 is equivalent to about 22% body fat, the minimal weight for height necessary for sustaining menstruation, and the 10th percentile at age 13 is equivalent to 17% body fat, the minimum for initiating menarche. A loss of body weight in the range of 10–15% of normal weight for height represents a loss of approximately one-third of the body fat, which will result in a drop below the 22% line and may result in abnormal menstrual function. 59

Although the nomogram is useful to show these relationships to patients, individual variation is such that the nomogram cannot be utilized to predict without fail the return of menses for an individual patient. Indeed, the accuracy of the nomogram has been challenged. 60 The fat criteria were derived from the indirect estimation of body fat from predicted total body water, with a regression equation that employs height and weight. There is no question that the most reliable and accurate method for estimating body fatness is hydrostatic weighing of body density, although dual-energy x-ray absorptiometry (DEXA) is also excellent. But one can hardly maintain a small pool for this purpose in a clinical office. Granted that the nomogram, and specifically the 22% body fat criterion, are not absolutely accurate; nevertheless, the concept is useful, and the nomogram remains helpful to illustrate the concept to patients.

The competitive female athlete has about 50% less body fat than the noncompetitor, very much under the 10th percentile for secondary amenorrhea (the 22% body fat line). This change in body fat can occur with no discernible change in total body weight, because fat is converted to lean muscle mass. 61 A critical look at the critical weight hypothesis argues that there is not a cause-and-effect relationship between body fat and menstrual function, only a correlation. 62 For this reason, considerable variation is seen with many examples of normal and abnormal menstrual function at all levels of body fat content. On the other hand, the correlation does exist, and body fat content and body weight are useful guides to the relationship between menstrual function and the energy balance of the body. Indeed, the leptin story (Chapter 19) has restored credibility to the critical weight hypothesis. It has always been 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.

In addition to the role of body fat, stress and energy expenditure appear to play an independent role. Warren has pointed out that dancers will have a return of menses during intervals of rest, despite no change in body weight or percent body fat. 63 High-energy output and stress, therefore, can act independently, as well as additively, to low body fat in suppressing reproductive function. It is not surprising that a woman with low body weight who is engaged in competitive activity (athletic or aesthetic) is highly susceptible to anovulation and amenorrhea.

Running in the dark is even more risky. Studies indicate that ovarian activity can be affected independently by strenuous activity and seasonal variation. 64 Decreased ovarian activity in autumn could be related to a greater dark photoperiod with increased pineal secretion of melatonin. Indeed, the conception rate of women living in northern Scandinavia is higher during the summer than in the winter. The practical conclusion is that serious runners can expect to encounter more problems with menstrual function in autumn and winter.

This menstrual disruption is similar to the hypothalamic dysfunction which is more marked in the classic cases of anorexia nervosa. Acute exercise decreases gonadotropins and increases prolactin, growth hormone, testosterone, ACTH, the adrenal steroids, and endorphins as a result of both enhanced secretion and reduced clearance. 65 The prolactin increase is in contrast to the absence of prolactin changes in undere nourished women. The prolactin increases are variable, small in amplitude, and exceedingly short in duration. Thus, it is unlikely that the prolactin increase is responsible for the suppression of the menstrual cycle. Most importantly, insignificant differences occur in prolactin when amenorrheic runners are compared with eumenorrheic runners or nonrunners. 66 In addition, women athletes have elevated daytime melatonin levels, and amenorrheic athletes have an exaggerated nocturnal secretion of melatonin. 67 The nocturnal increase in
Dieting due to neurotic fixation on food intake and weight.
Dieting for cosmetic reasons.
The anorectic reaction.

There are several important distinctions between the anorectic reaction and true anorexia nervosa. Psychologically, the patient with true anorexia nervosa has a...

Amenorrheic athletes are in a state of negative energy balance, further distinguished by elevated levels of insulin-like growth factor binding protein-1 (IGFBP-1), increased insulin sensitivity, decreased insulin levels, and increased growth hormone levels. Indeed, amenorrheic athletes who have cortisol levels that return to normal range regain menstrual function within 6 months, in contrast to athletes who maintain elevated cortisol levels and continue to be amenorrheic.

Amenorrheic athletes in a state of negative energy balance, further distinguished by elevated levels of insulin-like growth factor binding protein-1 (IGFBP-1), increased insulin sensitivity, decreased insulin levels, and increased growth hormone levels. The increase in IGFBP-1 can limit IGF activity in the hypothalamus and thus provide another mechanism for suppression of GnRH secretion.

The measurement of circulating b-endorphin levels may not reflect central mechanisms; both eumenorrheic and amenorrheic athletes have exercise-induced increases in the blood levels of b-endorphin.

Corticotropin-releasing hormone (CRH) directly inhibits hypothalamic 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), suggesting that this is the pathway by which stress interrupts reproductive function.

Amenorrheic athletes are prone to develop what can be called the anorectic reaction.

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.

The puzzle is why CRH is elevated in stress amenorrhea (especially that associated with weight loss) in contrast to 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 increases metabolism and weight loss.

Amenorrheic athletes do not have a diurnal pattern. 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 hypercortisolism. In addition, amenorrheic athletes have a blunted leptin response to the increase in insulin following meals. An increase in menstrual irregularity and amenorrhea is correlated with a decrease in body fat below 15% of body weight and leptin levels that are less than 3 ng/ml.

Because of the high levels of leptin present in overweight people, the purpose of 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 apparent resistance to leptin action associated with excess body weight and fat would then reflect resistance, but a lack of physiologic effect.

In regards to reproduction, the final pathway is suppression of GnRH, a response to multiple inputs indicating the availability of metabolic fuel. Even in runners with regular menstrual patterns, LH pulsatile frequency and amplitude are significantly reduced. A central inhibition of GnRH can be discerned even before there is perceptible evidence of menstrual irregularity. The clinical presentation (inadequate luteal phase, anovulation, or amenorrhea) will depend on the degree of GnRH suppression. Insufficient estrogen secretion at a critical time during growth can impede the growth spurt and yield short stature.

The characteristics of women in the subculture of exercise and amenorrhea strikingly remind one of anorexia nervosa: significant physical exercise, a necessity for control of the body, striving for artistic and technical proficiency, and the consequent preoccupation with the body, combined with the stressful pressures of performing and competition. Individuals in this lifestyle are prone to develop what can be called the anorectic reaction.

1. Dieting for cosmetic reasons.
2. Dieting due to neurotic fixation on food intake and weight.
3. The anorectic reaction.
4. True anorexia nervosa.

There are several important distinctions between the anorectic reaction and true anorexia nervosa. Psychologically, the patient with true anorexia nervosa has a misperception of reality and a lack of insight into the disease and her problem. She does not consider herself underweight and displays an impressive lack of concern.
over her dreadful physical condition and appearance. The clinician—patient relationship is difficult with no visible emotional involvement and a great deal of mistrust. Patients with the anorectic reaction have the capability for self-criticism. They can see the problem and describe it with insight and an absence of denial. The exercising woman and the competing athlete or dancer can develop an anorectic reaction. The anorectic reaction develops consciously and voluntarily, just as in anorexia nervosa, as the exercising woman deliberately makes an effort to decrease body weight. A clinician may be the first to be aware of the problem having encountered the patient because of the presenting complaint of either amenorrhea or now uncontrolled weight loss. Early recognition, concentrated counseling, and confidential support can intercept and prevent a progressive problem. The high prevalence of eating disorders among female athletes indicates that the threat of progression should not be underrated. It is estimated that the prevalence (in the general population, not just athletes) of the eating disorder is twice that of the full syndrome, and progression from pathologic dieting to the full syndrome can occur in some individuals in as short a time as 1–2 years.

Prognosis is excellent with early recognition, and simple weight gain may reverse the state of amenorrhea. The degree of reversibility is unknown, although general experience indicates that the majority of women regain ovulation when stress and exercise diminish or cease. However, these patients are often unwilling to give up their routines of exercise, and a sensitive clinician can perceive that the exercise is an important means for coping with daily life. Hormone therapy is, therefore, encouraged for these hypoestrogenic patients to provide protection against the loss of bone and cardiovascular changes. However, in patients with an eating disorder, restoration of normal hormonal levels is not sufficient to return bone density to normal; resumption of an adequate diet and weight gain are essential.

When pregnancy is desired, a reduction in the amount of exercise and a gain in weight should be recommended, or induction of ovulation must be pursued.

**Eating Disorders and Pregnancy**

It has been estimated that a typical pregnancy requires approximately 300 extra calories per day above that needed in the nonpregnant state. With sufficient caloric intake, weight gain during pregnancy averages 10–12 kg (22–26 lb). Women who are underweight prior to pregnancy need to increase their energy intake and gain 12–15 kg (26–33 lb). Imagine the reaction of a patient with an eating disorder when confronted with these facts. This is fuel for the fire (the morbid fear of fatness).

Prior to the 1970s, obstetricians vigorously advised patients to limit weight gain during pregnancy. This sadly misplaced advice can be traced to the false belief that excessive weight gain caused preeclampsia, made labor more difficult, and had a permanently detrimental impact on a woman’s figure. More appropriate recommendations began to emerge in the 1970s, based upon a growing body of scientific studies. These studies documented the importance of prepregnancy weight as well as weight gain during pregnancy as two very important determinants for infant birth weight.

The critical issue is the relationship between weight gain and pregnancy outcome. There are three classical studies of acute famine in Leningrad, Holland, and Wuppertal during the worst days of World War II. The mean birth weight during the siege of Leningrad declined 550 g to 2.789 g. During the Dutch famine, mean birth weight decreased by 300 g; there was no decrease in the birth weight of infants conceived during the famine whose mothers received adequate rations during the third trimester. In Wuppertal, mean birth weight was depressed by 170–227 g. These differences were proportional to the level of official rations, the conditions being the worst in Leningrad.

Studies of restriction of calories during pregnancy have indicated a ready achievement of lesser maternal weight gain, at the expense of lighter birth weights. Women who gain less than 20 pounds compared with those who gain more than 20 pounds are 2.3 times more likely to deliver infants of low birth weight and 1.5 times more likely to have a fetal death. These studies finally led to the abandonment of caloric limitation.

In general, studies of dietary supplementation have indicated increases in weight gains. The Special Supplemental Food Program for Women, Infants, and Children (WIC) was begun in the U.S. in 1973. A review of this program found a significant impact on decreasing preterm labor and delivery, an increase in mean birth weight, and a reduction in late fetal death. These improvements are attributed to an improved maternal physiologic status, not upgraded health care. The head circumferences of infants of mothers who participated in WIC were significantly larger, presumably reflecting accelerated brain growth.

Gradually it came to be appreciated that there is a linear relationship between birth weight and maternal weight gain at all levels of prepregnancy weight. However, prepregnancy weight increases, the importance of maternal weight gain diminishes. Thus, in underweight women, the importance of each factor, prepregnancy weight and the weight gain during pregnancy, is magnified.

Now that the many circumstances that influence infant birth weight are better recognized, a modern look at this subject is especially helpful. After adjusting for maternal age, race, parity, weight gain, socioeconomic status, cigarette consumption, and gestational age, there continues to be a statistically significant linear relationship between prepregnancy body mass and birth weight, as well as between prenatal weight gain and birth weight. Furthermore, the fetal death rate increases exponentially at lower weights. Most importantly, a low prepregnancy weight can be overcome; weight gain during pregnancy (beginning in the first trimester) in underweight women can bring an infant into the normal range for birth weight.

Low birth weight in an infant can be directly attributed mainly to two influencing factors: prematurity and fetal growth retardation. Underweight status before pregnancy and inadequate weight gain during the second half of pregnancy each separately increase the risk of preterm birth. In patients with eating disorders, the outcome of pregnancy is significantly influenced by both preterm delivery and intraterine growth retardation.

In view of the well-recognized correlation between maternal weight (prepregnancy weight and pregnancy weight gain) and infant size, it is certainly logical to expect a problem with pregnancy outcome in patients with eating disorders. Older reports were anecdotal in nature and usually failed to provide birth weights, because our awareness of the importance of body weight to pregnancy outcome is a relatively recent development.

More recent reports have documented problems with intraterine growth retardation and preterm labor. The average weight gain in 7 pregnancies in patients with anorexia nervosa was 8 kg. All infants demonstrated intraterine growth retardation in the third trimester, followed by accelerated growth after birth. A review of 23 pregnancies in 74 women treated for anorexia or bulimia documented the importance of the severity of the disorder. Women in remission gained more weight and had higher birth weights and 5-minute Apgar scores. Women with active disease had worsening symptoms and psychological problems during pregnancy. The smallest birth weights were born to those with anorexia and bulimia. All the women who were ill at conception continued to be ill. Nine of the 10 women who successfully breastfed for 6 months were in remission. The rate of preterm labor and delivery in patients with eating disorders is twice the normal incidence, and in a study of 50 Danish anorectic women, the perinatal mortality rate was 6 times the normal rate. In an Australian review of 14 consecutive women who failed to respond to clomiphene, all the women had a history of an eating or exercise disorder, and one-third of the eventual live births weighed less than 2500 g. There is a suggestion in retrospective data that the risk of spontaneous miscarriage is increased in pregnant women who are actively bulimic.

Some patients with eating disorders do well during pregnancy; however, even during pregnancy, many patients continue their abnormal eating behavior, and characteristically minimize or lie about their eating behaviors to clinicians, family, and friends. After the pregnancy, rapid deterioration usually takes place. In contrast, normal women are more accepting of their body size and weight gain during pregnancy, and dieting is unusual. Ultimately, the impact on weight is the key. There is no evidence that psychological distress independently influences fetal growth.

Expert help during the pregnancy is highly recommended. Some useful warning signs are the following:

1. Unusual concern regarding body shape.
2. Aversion to being weighed.
3. Lack of weight gain.

Approximately 25% of patients with anorexia are underweight due to self-imposed dietary restriction. Because underweight women frequently are anovulatory, it is not surprising that they represent a significant population in whom ovulation is induced. Comparing the outcome of pregnancy in underweight women after spontaneous and induced ovulation, it is apparent that inadequate weight has serious consequences. Underweight women who undergo induction of ovulation often fail to gain weight adequately during pregnancy despite care and counseling. As expected, intraterine growth retardation and preterm labor are significant problems.

The seriousness of the problem is heightened because of its presence at conception and during the first trimester. A study of adolescent pregnancy concluded that inadequate weight gain before 24 weeks gestation was associated with a significantly increased risk of having a small for gestational age infant, even when later...
weight gains brought the cumulative weight gain within normal adult standards. Later inadequate weight gain was associated with an increased risk of preterm delivery. These results suggest that prevention of preterm delivery and intrauterine growth retardation requires an effort encompassing the entire pregnancy, best beginning prior to conception. A rise in weight of more than 6 kg by the 28th week of pregnancy has predictive value, indicating with a high degree of probability that the rest of pregnancy will progress normally.

In summary, dietary restriction decreases birth weight and can have as great an impact as a serious famine. The impact on perinatal morbidity and mortality has been poorly assessed, but there is reason to believe that preterm labor is increased and that children with intrauterine growth retardation have more problems later in life. Although the impact of the eating disorder on parenting after the delivery has not been well studied, there is evidence that the children of parents with eating disorders are not adequately nourished. Mealtimes for these children are often filled with conflict, interfering with the important communication between parent and infant that usually takes place during feeding. The critical factor affecting the parental-infant interaction is the mother's preoccupation with her own body and shape.

Because intrauterine growth retardation before the third trimester is associated with significant long-term morbidity, clinicians treating amenorrhea associated with weight loss should consider solving the dietary problem before subjecting a fetus to a struggle during intrauterine life. Patients with an active eating disorder should wait for remission before getting pregnant. These messages, therefore, especially apply to two groups of physicians: obstetricians and reproductive endocrinologists.

When encountering a patient who is pregnant and has an eating disorder, the obstetrician should seek expert consultation in order to achieve and maintain remission of the eating disorder during the pregnancy. Careful monitoring of maternal weight gain and fetal growth is essential. Consideration should be given to special dietary supplementation, especially when delayed fetal growth is demonstrated. Full exploitation of the pregnancy is warranted to provide motivation for effective resolution of this psychodynamic disorder.

The best results can be achieved with stabilization of the disorder prior to pregnancy. The reproductive endocrinologist should hesitate before embarking on a program of ovulation induction. The reward of pregnancy should be offered as an inducement to reach a normal prepregnancy weight. Patients with an eating disorder who are considering pregnancy must be made aware of the potential adverse impact on fetal growth and development. The persistence of amenorrhea is not a good sign. Persistent amenorrhea is associated with longer duration of the eating disorder and more affective disorders.

The timing of psychiatric intervention is important. The prospect of pregnancy should stimulate a preconceptional effort by the patient's physicians. At no other time will the patient's physicians have such a strong ally, the motivating force behind the desire for pregnancy.

Inherited Genetic Defects

Specific inherited defects that cause hypogonadotropic hypogonadism have not been commonly recognized; however, with the increasing sophistication of molecular biology, this may change. No mutations of the a-subunit have been reported. Deficient secretion of GnRH is a consequence of Kallmann's syndrome and the inherited disorder of adrenal hypoplasia. A case of hypogonadism due to an LH b-subunit mutation and women with primary amenorrhea and delayed puberty due to a mutation in the FSH b-subunit have been reported. The mutated b-subunit genes produce alterations in the b-subunits that yield no immunoreactivity or bioactivity. Hence, hypogonadism will be associated with one high and one low gonadotropin level. Treatment with exogenous gonadotropins will achieve pregnancy in these rare patients; transplantation is autosomal recessive. When a high FSH and a normal or low LH are encountered, an elevated level of a-subunit and the presence of a pituitary mass indicate a gonadotroph adenoma.

Amenorrhea and Anosmia, Kallmann's Syndrome

A rare condition in females is the syndrome of congenital hypogonadotropic hypogonadism due to deficient secretion of GnRH, associated with anosmia or hyposmia, known as Kallmann's syndrome. There is a chronology of eponyms assigning credit for original descriptions of this syndrome, but with all due respect to the physicians who first recognized this association, it is far easier to remember it in a descriptive way, as a syndrome of amenorrhea and anosmia. The problem has a genetic and/or neurologic origin. In the former, this problem is seen as a failure in the migration of GnRH neurons, leading to the absence of the GnRH neurons. In the latter, this problem results from a failure in the migration of the olfactory neurons. This defect is a consequence of the failure of both olfactory axonal and GnRH neuronal migration from the olfactory placode in the forebrain. The cells that produce GnRH originate in the olfactory area and migrate during embryogenesis along fibers of the cranial nerve I complex (the terminal nerve) that connect the nose and the forebrain.

Three modes of transmission have been documented: X-linked, autosomal dominant, and autosomal recessive. The 5-7-fold increased frequency in males indicates that X-linked transmission is the most common. The Kallmann's syndrome can be associated with other disorders due to deletions or translocations of contiguous genes on the distal short arm of the X chromosome (such as X-linked short stature or ichthyosis and sulfatase deficiency).

The X-linked mutations (there is no consistent mutation) responsible for this syndrome involve a single gene (KAL) on the short arm of the X chromosome that encodes a protein responsible for functions necessary for neuronal migration. This protein has been named anosmin-1. The syndrome of anosmia and amenorrhea associated with X-linked mutations results from the failure of this nerve complex to penetrate the forebrain, preventing the successful migration of GnRH neurons. Other neurologic abnormalities (optic atrophy, lacrimal deficiency, cleft lip and palate, and cleft palate) can be present, suggesting more widespread neurologic defects. Renal and bone abnormalities, hearing deficit, color blindness, and cleft lip and palate (the most common associated abnormality) also occur in affected individuals, probably reflecting the fact that the gene is expressed in tissues other than the hypothalamus. The syndrome occurs as an inherited or sporadic defect.

Molecular Explanations for Hypogonadotropic Amenorrhea

It is possible that patients with hypothalamic amenorrhea due to isolated deficiency of GnRH secretion (and no other abnormalities) have a defect similar to that of Kallmann's syndrome. With a lesser penetrance, only the GnRH migratory defect is expressed. In some individuals with amenorrhea and a normal sense of smell, family members can be identified with anosmia. Some patients with GnRH deficiency, but without anosmia, have an autosomal mode of transmission; however, a defect in the GnRH gene has not been commonly detected and X-linked mutations are uncommon.

A brother and sister have been described with hypogonadotropic hypogonadism due to mutations of the GnRH receptor gene. The parents and one sister were heterozygous and normal; thus, the mutations were transmitted as an autosomal recessive trait. Screening 46 men and women with hypogonadotropic hypogonadism revealed the existence of an autosomally inherited GnRH gene receptor mutation in one of the 14 families with an affected female. This mutation produces an interference in signal transduction resulting in resistance to GnRH stimulation. Thus this interferon mutation causes hypogonadotropic amenorrhea that is not readily treated with GnRH, but response to exogenous gonadotropins is unimpaired. Because most individuals with hypogonadotropic amenorrhea respond to GnRH, this genetic defect is not common, and worth pursuing only in patients who have family members with similar presentations. This approach is supported by the failure to detect GnRH receptor gene deletions or rearrangements in patients with idiopathic hypogonadotropic hypogonadism.

Adrenal Hypoplasia

Adrenal hypoplasia is an X-linked inherited disorder that results in adrenal insufficiency, and in survivors, hypogonadotropic hypogonadism. This condition is caused by mutations in the DAX-1 gene, a gene that encodes a protein that is similar in structure to receptors that have no identified ligands (orphan receptors). These rare patients have both deficient secretion of GnRH and impaired response to GnRH.

Postpill Amenorrhea

In the past, it was assumed that secondary amenorrhea reflected persistent suppressive effects of oral contraceptive medication or the use of the intramuscular depot form of medroxyprogesterone acetate (Depo-Provera). It is now recognized that the fertility rate is normal following discontinuance of either of these forms of contraception (Chapter 22), and attempts to identify a cause-effect relationship in case-control studies have failed. Therefore, amenorrhea following the use of
steroids for contraception requires investigation as described in order to avoid missing a significant problem. This investigation should be pursued if a patient is amenorrheic 6 months after discontinuing oral contraception or 12 months after the last injection of Depo-Provera.

Hormone Therapy

The patient who is hypoestrogenic and who is not a candidate for induction of ovulation deserves hormone therapy. This includes patients appropriately evaluated and diagnosed as having gonadal failure, patients with hypothyroidic amenorrhea, and postgonadectomy patients. The long-term impact of the hypoestrogenic state in terms of cardiovascular disease has long been recognized. To some degree, the beneficial impact of exercise on the lipoprotein profile is reversed by estrogen deficiency; however, athletes have high, exercise-induced levels of HDL-cholesterol, which should be cardioprotective. 325 326

We want to emphasize that the bone density in women is dependent on normal reproductive age levels of estrogen and progesterone. Even the most strenuous of exercise does not balance the consequences of hypoestrogenism on the bones, especially in adolescents. 327 328 329 330 331 332 The degree of bone loss is influenced by the specific athletic activity; e.g., gymnasts have a higher bone density than runners despite similar menstrual patterns and body fat. 81 In one study, ballet dancers were able to maintain bone density at weight-bearing sites, despite oligomenorrhea and reduced body weight, whereas another study found reduced bone mass in weight-bearing bones. 333 334 It makes sense that different exercises have different osteogenic effects according to the mechanical forces generated. In addition, the effect of bone loss is greater in the spine because trabecular bone is more sensitive to the loss of estrogen. Whether specific individuals and activities are threatened with osteoporosis-related fractures later in life requires follow-up data that are not available at this point in time.

In the absence of estrogen, the normal response of bone to stress (to become stronger) is impaired. 335 The same arguments that apply to hormone treatment in older women (Chapter 17 and Chapter 18) can be convincingly used to encourage these younger women to replace the estrogen they are lacking. The amenorrheic exerciser should be made aware that the hypoestrogenic state is associated with a greater risk of stress fractures. 336 337 338 339 340 341 342 Indeed, the loss of bone in amenorrheic athletes, with some exceptions as noted above, occurs at all skeletal sites that are weight bearing and subject to stress fractures. 343 Ballet dancers with delayed menarche are more prone to scoliosis as well as stress fractures. 344 It is not certain, however, whether this greater risk of stress fractures is influenced solely by bone density changes that some studies fail to correlate fractures with reductions in bone density. 345 It should be noted that bone loss in amenorrheic women shows the same pattern over time as seen in postmenopausal women. 346 The loss is most rapid in the first few years, emphasizing the need for early treatment. The bone density increase in response to estrogen-progestin treatment of women with exercise-induced amenorrhea is impressive and worth achieving. 347 Long-term estrogen treatment of women with Turner syndrome effectively maintains bone density. 348

It is worth emphasizing the importance of bone mineralization early in life. The subsequent risk of fracture from osteoporosis will depend upon bone mass at the time of menopause. The rate of bone loss following menopause. There is only a relatively narrow window of opportunity for acquiring bone mass. Almost all of the bone mass in the hip and the vertebral bodies will be accumulated by late adolescence (age 18), and the years following menarche (11–14) are especially important. 349 350

The importance of a normal diet and normal hormonal support during adolescence cannot be overlooked.

In patients with eating disorders, bone density correlates with body weight. 351 352 The response to hormone therapy will be impaired as long as an abnormal weight is maintained. 353 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. 354 355 Interventions that improve diet and reduce overtraining can restore hormonal function, and even improve athletic performance (presumably because of better energy balance). 356

Several reports have indicated that patients with hyperprolactinemia are at risk for osteoporosis. At first this appeared not to be related to estrogen status, suggesting an independent effect of prolactin. Results have been confusing for several reasons. Controls in various studies were matched in different fashions; e.g., only for age, fertility status, and body weight. Other studies have matched the method of study in some reports, has a reduced sensitivity and significant variation when used to assess the response of the axial skeleton. And finally, the estrogen status of the hyperprolactinemic patients was not always carefully quantified. It is now recognized that the bone density changes observed in hyperprolactinemic amenorrheic women are due to the hypoestrogenic state. 357 358 359

The standard program for estrogen treatment should be used. A good schedule is the following: 0.625 mg conjugated estrogens or 1 mg estradiol daily with 5 mg medroxyprogesterone acetate for 2 weeks every month. If the postmenopausal agent is responsible for side effects, patients do well upon changing to 0.7 mg norethindrone. In a few individuals, the estrogen dosage may have to be increased in order to achieve menstrual bleeding. Whether a flow-provoking dose of estrogen is necessary for optimal protection of the bones has not been addressed in a clinical study. In patients who have not undergone pubertal development, a lower dose regimen should be used initially, as outlined in Chapter 10. Periodic measurements of bone density are worthwhile to assess adequacy of hormonal treatment and to provide evidence of lifestyle and dieting changes.

Menstruation generally occurs 3 days after the last day of progesterin medication. Bleeding that occurs at any time other than the usual expected time may be a sign that endogenous function has returned. The hormone treatment program should be discontinued and the patient monitored for the resumption of ovulation.

The importance of menstrual irregularities in young women cannot be overemphasized. Regular and visible menstrual bleeding is often a gratifying experience in the young patient with gonadal dygenesis and one that reinforces her identification with the feminine gender role. On the other hand, serious exercisers (such as athletes and dancers) may want to avoid menstrual bleeding. One can provide hormone therapy to these women utilizing the daily combination approach: 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate or equivalent preparations given together every day without a break.

If for some reason, a hypoestrogenic woman refuses hormone treatment, supplemental calcium (1000–1500 mg daily) should be strongly encouraged. High calcium intake when combined with a high level of exercise is more effective in protecting the vertebral bone density than either exercise or calcium alone. 357 Even patients receiving hormone therapy should be cautioned to maintain a normal intake of calcium and vitamin D, and this usually requires supplementing the teenage diet.

Endogenous estrogen is necessary for optimal protection of the bones has not been addressed in a clinical study. In patients who have not undergone pubertal development, a lower dose regimen should be used initially, as outlined in Chapter 10. Periodic measurements of bone density are worthwhile to assess adequacy of hormonal treatment and to provide evidence of lifestyle and dieting changes.

Menstruation generally occurs 3 days after the last day of progesterin medication. Bleeding that occurs at any time other than the usual expected time may be a sign that endogenous function has returned. The hormone treatment program should be discontinued and the patient monitored for the resumption of ovulation.

The importance of menstrual irregularities in young women cannot be overemphasized. Regular and visible menstrual bleeding is often a gratifying experience in the young patient with gonadal dygenesis and one that reinforces her identification with the feminine gender role. On the other hand, serious exercisers (such as athletes and dancers) may want to avoid menstrual bleeding. One can provide hormone therapy to these women utilizing the daily combination approach: 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate or equivalent preparations given together every day without a break.

If for some reason, a hypoestrogenic woman refuses hormone treatment, supplemental calcium (1000–1500 mg daily) should be strongly encouraged. High calcium intake when combined with a high level of exercise is more effective in protecting the vertebral bone density than either exercise or calcium alone. 357 Even patients receiving hormone therapy should be cautioned to maintain a normal intake of calcium and vitamin D, and this usually requires supplementing the teenage diet.

Patients with hypothalamic amenorrhea must be cautioned that hormone therapy will not protect against pregnancy in the event that normal function unknowingly returns. In the patient who desires effective contraception, it is reasonable to utilize a low-dose oral contraceptive to provide the missing estrogen. 358 This is an excellent option in patients with premature ovarian failure because a spontaneous resumption of ovarian function can occur without warning. Athletes and others interested in avoiding menstruation may take oral contraceptives every day without a pill-free interval.

It is not enough to provide hormone therapy when disturbed menstrual function is secondary to psychobiologic stress responses. Appropriate support and counseling are necessary to help patients develop coping mechanisms other than extreme dieting and exercise. All available skills and resources should be utilized to promote healthy attitudes and healthy behaviors. The presence of amenorrhea in athletes and recreational exercisers should be regarded as a sign of negative energy balance, a condition requiring appropriate interventions.

Chapter References

5. Sarlis NJ, Bruucker-Davis F, Dorman JL, Skarulis MC, MRI-demonstrable regression of a pituitary mass in a case of primary hypothyroidism after a week of acute thyroid hormone therapy, J Clin Endocrinol Metab 82:808, 1997.
Pathogenesis of Anovulation
Central Defects
Abnormal Feedback Signals
Luteal Ovarian Cortisol
Estradiol Body Weight
Precocious Puberty
The Polycystic Ovary
Gonadotropin-Stimulating Hormone
P450c17 Dysregulation
Insulin Resistance, Hyperinsulinemia, and Hyperandrogenism
How Does Hyperinsulinemia Produce Hyperandrogenism?
Can Ovarian Hyperandrogenism be Treated With Diabetes Drugs?
The Clinical Consequences of Persistent Anovulation
Who Should Be Tested for Hyperinsulinemia?

Conclusion

Chapter References

Pathogenesis of Anovulation

Just before and during menses, escape from the negative feedback of estrogen, progesterone, and inhibin results in increased follicle-stimulating hormone (FSH) secretion by the anterior pituitary. This initial increase in FSH is essential for follicular growth and steroidogenesis. With continued growth of the follicle, autocrine/paracrine factors produced within the follicle maintain follicular sensitivity to FSH, allowing conversion from a microenvironment dominated by androgens to one dominated by estrogen, a change necessary for a complete and successful follicular lifespan. Continuing and combined action of FSH and activin leads to the appearance of luteinizing hormone (LH) receptors on the granulosa cells, prerequisite for ovulation and lutetization. Ovulation is triggered by the rapid rise in circulating levels of estradiol. A positive feedback response at the level of the anterior pituitary (and perhaps at the hypothalamus as well) results in the midcycle surge of LH necessary for expulsion of the egg and formation of the corpus luteum. A rise in progesterone follows ovulation along with a second rise in estradiol, producing the 14-day luteal phase characterized by low FSH and LH levels. The demise of the corpus luteum, coincident with a fall in hormone levels, allows FSH to increase again, thus initiating a new cycle.

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 thecal androgen synthesis. In the late follicular phase, increased production of inhibin by the granulosa (and decreased activin) promotes androgen synthesis in the theca in response to LH and insulin-like growth factor-II (IGF-II) to provide substrate for even greater estrogen production in the granulosa. 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. In the mature granulosa, activin serves to prevent premature lutetization 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 and growth factor expression. The activin proteins (which enhance FSH activity) are produced in greatest amounts early in follicular development to enhance follicle receptivity to FSH.

The right concentration of androgens within granulosa cells promotes aromatase activity and inhibin production, and, in turn, inhibin promotes LH stimulation of thecal androgen synthesis. LH stimulation of androgen production in theca cells is further enhanced by the autocrine activity of insulin-like growth factor. Evidence indicates that the endogenous insulin-like growth factor in the human ovarian follicle is IGF-II in both the granulosa and the thecal cells. 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. With development of the follicle, inhibin expression 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.

This recycling mechanism is regulated by substances functioning as classic hormones (FSH, LH, estradiol, and inhibin) transmitting messages between the ovary and the hypothalamic-pituitary axis, and autocrine/paracrine factors (IGF-II, inhibin, and activin, among others), which coordinate sequential activities within the follicle destined to ovulate. The negative feedback relationship between corpus luteum products (estradiol, progesterone, and inhibin) and FSH results in the critical initial rise in FSH during menses, and the positive feedback relationship between estradiol and LH is the ovulatory stimulus. Within the ovary, IGF-II, inhibin, and activin modify follicular receptor responses necessary for growth and function. Dysfunction in the cycle can be due to an abnormality in one of the various roles for any one of these substances or an inability to respond to signals.

Central Defects

The hypothalamic-pituitary axis may be unable to respond, even if given adequate and appropriately timed feedback signals. A pituitary tumor represents an obvious example of a central defect in menstrual function and is discussed in Chapter 11. Amenorrhea.

Although difficult to demonstrate definitively, malfunction within the hypothalamus is both a likely, as well as a favorite, explanation for ovulatory failure. Normal pituitary ovulatory response to the follicle's steroid signals requires the presence of gonadotropin-releasing hormone (GnRH) pulsatile secretion within a critical range. The teenager between menarche and the onset of ovulation cannot generate a normal cycle until full GnRH pulsatile secretion is achieved. Increasing intensity of GnRH suppression is associated with increasing dysfunction and a changing clinical presentation. A variety of problems, such as stress and anxiety, borderline anorexia nervosa, and acute weight loss after a crash diet, is associated with an inhibition of normal GnRH pulsatile secretion. The mechanism for this suppression of GnRH is excessive hypothalamic activity of corticotropin-releasing hormone (CRH), a response to stress. Patients present more commonly with amenorrhea, as discussed in Chapter 11. However if GnRH is only partially suppressed, homeostatic pituitary-ovarian function is maintained, and the patients will be ovulatory. 1 2 3 and 5

12Anovulation and The Polycystic Ovary
At least one specific clinical syndrome of central anovulatory dysfunction has been recognized: hyperprolactinemia. Increasing levels of prolactin can cause a woman to progress through a spectrum, beginning with an inadequate luteal phase to anovulation to the amenorrhea associated with complete GnRH suppression. A search for galactorrhea and measurement of the prolactin level are important screening procedures for all women who are not ovulating normally. The presence of galactorrhea or elevated prolactin levels dictates a choice of dopamine agonist treatment for the induction of ovulation. Indeed, even in women with normal prolactin levels, long-term treatment with a dopamine agonist can reduce androgen levels and restore menstrual function, either through a direct central dopaminergic effect or because of further suppression of normal prolactin levels.

Anovulatory women with polycystic ovaries have a higher LH (and presumably GnRH) pulse frequency and amplitude when compared to the normal midfollicular phase. Central opioid tone appears to be suppressed because there is no difference in response to naloxone. The enhanced pulsatile secretion of GnRH can be attributed to a reduction in hypothalamic opioid inhibition because of the chronic absence of progesterone. Interaction at the dopamine-endorphin sites may be altered because pretreatment with a dopamine precursor leads to a naloxone-induced increase in LH in anovulatory women compared with controls.

Abnormal Feedback Signals

Abnormal cycles can be due to failures within the system or due to the introduction of confounding factors. In order to achieve the appropriate changes within the cycle, estradiol levels must rise and fall in synchrony with morphologic events. Therefore, two possible signal failures may occur: 1) estradiol levels may not fall low enough to allow sufficient FSH response for the initial growth stimulus, and 2) levels of estradiol may be inadequate to produce the positive stimulatory effects necessary to induce the ovulatory surge of LH.

Loss of FSH Stimulation

In order to achieve recycling, a nadir in blood sex steroid levels must occur so that the initial event in the cycle, the rise in FSH, can take place. Sustained estrogen at such a key moment would not permit FSH stimulation of follicular growth and maturation, and recycling would be thwarted. The necessary decline in blood estrogen requires reduction of secretion, appropriate clearance and metabolism, and the absence of a significant contribution of estrogen to the circulation by extragonadal sources.

Persistent Estrogen Secretion. The most common clinical example of anovulation associated with continued secretion of sex steroids is pregnancy. Persistent and elevated secretion of estrogen can be encountered rarely with an ovarian or adrenal tumor. In such a case, anovulation or amenorrhea may bring the patient to a clinician's attention.

Abnormal Estrogen Clearance and Metabolism. The clearance and metabolism of estrogen can be impaired by other pathologic conditions, such as thyroid or hepatic disease. It is for this reason that a careful history and physical examination are important elements in the differential diagnosis of anovulation. Both hyperthyroidism and hypothyroidism can cause persistent anovulation by altering not only metabolic clearance but also the peripheral conversion rates among the various steroids. The subtle presence of hypothyroidism, which may be associated with elevated prolactin levels, demands screening of anovulatory and amenorrheic women with a thyroid-stimulating hormone (TSH) level.

Extraglandular Estrogen Production. Extragonadal contribution to the blood estrogen level can reach significant proportions. Although the adrenal gland does not secrete appreciable amounts of estrogen into the circulation, it indirectly contributes to the total estrogen level. This is accomplished by the extraglandular peripheral conversion of C-19 androgenic precursors, mainly androstenedione, to estrogen. In this manner, psychological or physical stress may increase the adrenal contribution of estrogenic precursor, and subsequent conversion to estrogen may sustain the blood level of estrogen at a time when a decline is necessary for successful recycling of the menstrual cycle. Adipose tissue is capable of converting androstenedione to estrogen; hence, the percent conversion increases with increasing body weight. This is at least one mechanism for the well-known association between obesity and anovulation.

Loss of LH Stimulation

A failure in gonadal production of estrogen need not be absolute. Obviously, the patient with gonadal dysgenesis and ovarian failure will present with amenorrhea and infertility because of a total lack of estrogen secretion. More commonly, the clinician is concerned with the patient who has gonadotropin and estrogen production but does not ovulate. The failure to achieve a critical midcycle level of estradiol necessary to trigger the gonadotropin surge may be due to a relative deficiency in steroid production. The perimenopausal woman undergoes a terminal period of anovulation, which may represent steroidogenic refractoriness within the remaining elderly follicles. This inadequacy may be due to intrinsic follicular weaknesses or impairment in the follicular-gonadotropin interaction. In any case, the end result is the same—a failure to achieve critical signal levels of estradiol at the appropriate time in midcycle.

Local Ovarian Conditions

An understanding of the critical balances within the follicle indicates possible points of failure that may lead to anovulation. Local autocrine/paracrine factors prevent atresia despite declining FSH levels by enhancing the action of FSH in increasing the number of FSH receptors within the follicle, thus increasing follicular sensitivity to FSH. In addition, these factors enhance the induction of LH receptors by FSH, making it possible for the follicle to respond to the LH surge at midcycle. A follicle can fail to grow and ovulate because of inadequate expression or impaired function of any of the following local ovarian activities (described in detail in Chapter 6):

1. 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.
2. Derived from the dominant follicle, estradiol levels increase steadily and, through negative feedback effects, exert a progressively greater suppressive influence on FSH release.
3. Insulin-like growth factor-II (IGF-II) is produced in theca cells in response to gonadotropin stimulation, and this response is enhanced by estradiol and growth hormone. In an autocrine action, IGF-II increases LH stimulation of androgen production in theca cells.
4. IGF-II stimulates granulosa cell proliferation, aromatase activity, and progestosterone synthesis.
5. FSH inhibits IGF binding protein synthesis and thus maximizes growth factor availability.
6. FSH stimulates inhibin and activin production by granulosa cells.
7. Activin augments FSH activities: FSH receptor expression, aromatization, inhibin/activin production, and LH receptor expression.
8. Inhibin enhances LH stimulation of androgen synthesis in the theca to provide substrate for aromatization to estrogen in the granulosa.
9. While directing a decline in FSH levels, the midfollicular rise in estradiol exerts a positive feedback influence on LH secretion. LH levels rise steadily during the late follicular phase, stimulating androgen production in the theca.
10. The positive action of estrogen also includes modification of the gonadotropin molecule, increasing the quality (the bioactivity) as well as the quantity of LH at midcycle.
11. Inhibin and, less importantly, follistatin, secreted by the granulosa cells in response to FSH, directly suppress pituitary FSH secretion.
12. FSH induces the appearance of LH receptors on granulosa cells.

The factors that control follicular growth and development are now understood in terms of the two-cell explanation described in Chapter 2 and Chapter 6. A very precise coordination is necessary between morphologic development and hormone stimulation. Perturbations may arise from an infectious process, from the presence of endometriosis, or by abnormal qualitative or quantitative changes in tropic hormone receptors (ovarian insensitivity), or the necessary biologic effects may be blocked by an improper molecular constitution of the gonadotropins (heterogeneity of the glycopeptide hormones).
Increased peripheral aromatization of androgens to estrogens. Decreased levels of sex hormone-binding globulin (SHBG), resulting in increased levels of free estradiol and testosterone.

12

Ovarian Failure. Hypergonadotropic hypogonadism, the inability of the ovary to respond to any gonadotropic stimulation, usually due to the absence of follicular development. Alternately, it can result from a central or local defect. The polycystic ovary syndrome (PCOS) can reflect any of the dysfunctional states discussed above. In other words, the polycystic ovary is the result of a functional derangement, not a specific structural defect.

Anovulatory Dysfunction. The patient who has asynchronous gonadotropin and estrogen production and does not ovulate presents with a variety of clinical symptoms. Some patients may have amenorrhea, hirsutism, and obesity, together with a demonstration of enlarged, polycystic ovaries.

The frequency of obesity in women with anovulation and polycystic ovaries has been reported to be from 35% to 60%. Obesity is associated with three alterations that interfere with normal ovulation, and weight loss improves all three:

1. Increased peripheral aromatization of androgens to estrogens.
2. Decreased levels of sex hormone-binding globulin (SHBG), resulting in increased levels of free estradiol and testosterone.
3. Increased insulin levels that can stimulate ovarian stromal tissue production of androgens.

Precise Etiology

The normal ovulatory function of the menstrual system relies on a dynamic coordination of complex actions. Abnormal function may represent discordance at all of the levels reviewed in the above paragraphs. Thus, a minor deficiency in the estradiol signal will be associated with a subnormal central response and an impaired or inappropriate degree of follicular growth and function. Dysfunction is sustained by the internal feedback mechanisms within the system, and anovulation can become a persistent problem.

It is usually impossible to reduce the issue of etiology to a single factor of abnormal menstrual function, except in severe states such as pituitary tumors, anorexia nervosa, gonadal dysgenesis, and perhaps hyperprolactinemia and obesity. Not only is it often impossible, but it is usually unnecessary to define the precise etiology. Regardless of the nature of the initial cause of the problem, the final clinical statement of the dysfunction is predictable, and easily diagnosed and managed. In patients who have abnormal or absent menstrual function, but are otherwise medically normal, the diagnosis will fall into one of three categories:

1. Ovarian Failure. Hypogonadotrophic hypogonadism, the inability of the ovary to respond to any gonadotropic stimulation, usually due to the absence of follicular tissue on a genetic basis (discussed in Chapter 11).
2. Central Failure. Hypogonadotrophic hypogonadism, hypothalamic or pituitary suppression as expressed in abnormal low or normal serum gonadotropins (discussed in Chapter 11).
3. Anovulatory Dysfunction. The patient who has asynchronous gonadotropin and estrogen production and does not ovulate presents with a variety of clinical manifestations. The associated clinical signs and symptoms depend on the level of gonadal function preserved and are represented by the following principal problems:
   - Amenorrhea (Chapter 11).
   - Hirsutism (Chapter 13).
   - Dysfunctional uterine bleeding (Chapter 15).
   - Endometrial hyperplasia and cancer (Chapter 18).
   - Breast disease (Chapter 16).
   - Infertility and induction of ovulation (Chapter 30).

The Polycystic Ovary

In 1935, Irving F. Stein and Michael L. Leventhal first described a symptom complex associated with anovulation. Both gynecologists were born in Chicago, both were graduates of Rush Medical College, and both spent their entire professional careers at Michael Reese Hospital. Stein and Leventhal described 7 patients (4 of whom were obese) with amenorrhea, hirsutism, and enlarged, polycystic ovaries. They reported the results of bilateral wedge resection, removing one-half to three-fourths of each ovary; all 7 patients resumed regular menses, and 2 became pregnant. Stein and Leventhal developed the wedge resection after they observed that several of their amenorrheic patients menstruated after ovarian biopsies. They reasoned that the thickened tunic was preventing follicles from reaching the surface of the ovary.

Acceptance of this syndrome as a singular clinical entity led to a rather rigid approach to this problem for many years. Only those women qualified who had a history of oligomenorrhea, hirsutism, and obesity, together with a demonstration of enlarged, polycystic ovaries, a clinical state now recognized to be characteristic of extreme cases. It is far more useful clinically to avoid the use of eponyms and even the term polycystic ovary syndrome or disease. It is better to consider this problem as one of persistent anovulation with a spectrum of etiologies and clinical manifestations, that now includes insulin resistance and hyperinsulinemia, as well as hyperandrogenism. Of course, specific conditions must be pursued and excluded, such as adrenal hyperplasia, Cushing’s syndrome, hyperprolactinemia, and androgen-producing tumors.

A question that has puzzled gynecologists and endocrinologists for many years is what causes polycystic ovaries. There is an answer that is appealing in its logic and clinical applicability. The characteristic polycystic ovary emerges when a state of anovulation persists for any length of time. Whether diagnosis is by ultrasonography or by the traditional clinical and biochemical criteria, a cross section of anovulatory women at any one point of time will reveal that approximately 75% will have polycystic ovaries. Because there are many causes of anovulation, there are many causes of polycystic ovaries. A similar clinical picture and ovarian condition can reflect any of the dysfunctional states discussed above. In other words, the polycystic ovary is the result of a functional derangement, not a specific central or local defect.
Insistence on a specific endocrine or clinical criterion for the diagnosis of the polycystic ovary syndrome results in the inclusion of a collection of patients that represents a focused segment isolated from the broad clinical spectrum in which these patients really belong. This especially applies to the use of ultrasonography to make the diagnosis of the polycystic ovary syndrome (the presence of an increased number of ovarian follicles, often in a necklace-like pattern, and an increase in ovarian volume, largely due to an increase in stroma). From 8% to 25% of normal women will demonstrate ultrasonographic findings typical of polycystic ovaries. Even 14% of women on oral contraceptives have been found to have this ultrasonographic picture. Ultrasonography as a diagnostic tool for this condition is unnecessary, and we vigorously discourage its use for this purpose. Magnetic resonance studies further confirm the unreliability of the imaging finding presumed to be diagnostic of this condition.

It is argued that normally ovulating women with polycystic ovaries on ultrasonography have underlying metabolic abnormalities. However, the great majority of ovulatory women with polycystic ovaries on ultrasonography are endocrinologically normal, and only occasionally is an androgen level minimally elevated. We are not convinced that such subtle changes are of clinical importance. Even if this argument were correct and a basic disorder is present, but homeostatic adjustments allow the maintenance of normal physiologic mechanisms and there are no clinical consequences, it is hard to justify medical interventions.

In contrast to the characteristic picture of fluctuating hormone levels in the normal cycle, a "steady state" of gonadotropins and sex steroids can be depicted in association with persistent anovulation. This steady state is only relative, and is being exaggerated in the diagram to present a concept of this clinical problem. In patients with persistent anovulation, the average daily production of estrogen and androgens is both increased and dependent on LH stimulation. This is reflected in higher circulating levels of testosterone, androstenedione, dehydroepiandrosterone (DHA), dehydroepiandrosterone sulfate (DHAS), 17-hydroxyprogesterone (17-OHP), and estrone.

The ovary does not secrete increased amounts of estrogen, and estradiol levels are equivalent to early follicular phase concentrations. Circulating estrone levels are slightly elevated. The increased total estrogen is due to peripheral conversion of the increased amounts of androstenedione to estrone. That is not to say that there is no ovarian secretion of estrogen. Both estrogen and estradiol continue to be secreted in significant, although low, amounts.

When compared with levels found in normal women, patients with persistent anovulation have higher mean concentrations of LH, but low or normal levels of FSH.

The elevated LH levels are partly due to an increased sensitivity of the pituitary to releasing hormone stimulation, manifested by an increase in LH pulse amplitude and frequency, but mainly amplitude. This is consistent with the concepts discussed in Chapter 5, linking a high estrogen environment with anterior pituitary secretion of LH and suppression of FSH. It is noteworthy that this high level of LH is characterized by an increased level of LH bioactivity.

The gonadotropin pattern (high LH and low FSH) can also be due to increased frequency of GnRH pulsatile secretion. Central opioid tone appears to be suppressed because there is no difference in response to naloxone. Indeed, the enhanced pulsatile secretion of GnRH can be attributed to a reduction in hypothalamic opioid inhibition because of the chronic absence of progesterone. This is associated with an increase in amplitude and frequency of LH secretion that is correlated with the level of circulating estrogen. It is likely that this increased activity is taking place at both hypothalamic and pituitary sites. This altered state is also associated with a change in the circadian pattern with the highest LH values occurring in late afternoon rather than at night.

The increase in LH pulse frequency and pituitary secretion of LH are characteristic of the anovulatory state, and are independent of obesity. Obesity attenuates the LH response to GnRH, and LH pulse amplitude is relatively normal in overweight women with polycystic ovaries, although the increase in pulse frequency is maintained. The overall LH reduction may be a direct effect of the greater hyperandrogenism in obese women.

The increased pituitary and hypothalamic sensitivity can be attributed to the increased estrone levels, but an important contributing factor is the impact of the decreased SHBG concentration. The levels of sex hormone-binding globulin (SHBG) are controlled by a balance of hormonal influences on its synthesis in the liver; testosterone is inhibitory, estrogen and thyroxine are stimulatory. In anovulatory women with polycystic ovaries, there is an approximately 50% reduction in circulating levels of SHBG, a response to the increased testosterone, and in patients with hyperinsulinemia, due to a direct insulin effect on the liver (discussed later in this chapter). Despite no increase in estradiol secretion, free estradiol levels are increased because of the significant decrease in SHBG. The clinical consequences of uninterrupted estrogen stimulation (endometrial and perhaps breast cancer) as well as the increased LH are the result of the two estrogenic influences, estrone and free estradiol.

The increased LH secretion as expressed by the LH:FSH ratio is positively correlated with the increased free estradiol. The lower FSH levels represent the sensitivity of the FSH negative feedback system to the elevated estrogen, both free estradiol and the estrone formed from peripheral conversion of androstenedione. In addition, the altered pattern of GnRH secretion can contribute to this characteristic LH:FSH ratio. There is no evidence to support a role for inhibin suppression of FSH. In fact, consistent with impaired granulosa cell function, inhibin production in granulosa cells from polycystic ovaries is reduced. However, it is possible that these early findings with inhibin reflect the lack of precision associated with the first immunoassays. A sensitive assay for inhibin B has detected high levels in women with polycystic ovaries, suggesting that multiple small follicles can suppress FSH levels by increasing the circulating levels of inhibin B.

Because the FSH levels are not totally depressed, new follicular growth is continuously stimulated, but not to the point of full maturation and ovulation. Despite the fact that full growth potential is not realized, follicular lifespan may extend several months in the form of multiple follicular cysts, 2–10 mm in diameter (some can be as large as 15 mm). These follicles are surrounded by hyperplastic theca cells, often luteinized in response to the high LH levels. The accumulation of follicular tissue in various stages of development allows an increased and relatively constant production of steroids in response to the gonadotropin stimulation. This condition is self-sustaining. As various follicles undergo atresia, they are immediately replaced by new follicles of similar limited growth potential.

The tissue derived from follicular atresia is also sustained by the steady state and now contributes to the stromal compartment of the ovary. In terms of the two-cell explanation of follicular steroidogenesis, atresia is associated with a degenerating granulosa, leaving the theca cells to contribute to the stromal compartment of the ovary. It is not surprising, therefore, that this functioning stromal tissue secretes significant amounts of androstenedione and testosterone, the usual products of theca
There are 4 times more ovarian hilus cell nests (hyperplasia).

The surface area is doubled, giving an average volume increase of 2.8 times. The same number of primordial follicles is present, but the number of growing and atretic follicles is doubled. Each ovary may contain 20–100 cystic follicles.

In women with hirsutism, the steroidogenic response to ACTH stimulation was not consistent with an inherent disorder in P450c17. Adrenal product, testify to adrenal participation. The mechanism and the clinical importance of this involvement will be further discussed in

P450c17 Dysregulation

The polycystic ovary is usually enlarged and is characterized by a smooth pearly white capsule. For years, it was erroneously believed that the thick sclerotic capsule acted as a mechanical barrier to ovulation. A more accurate concept is that the polycystic ovary is a consequence of the loss of ovulation and the achievement of the steady state of persistent anovulation. The characteristics of the ovary reflect this dysfunctional state:

1. The surface area is doubled, giving an average volume increase of 2.8 times.
2. The same number of primordial follicles is present, but the number of growing and atretic follicles is doubled. Each ovary may contain 20–100 cystic follicles.
3. The thickness of the tunica (outermost layer) is increased by 50%.
4. A one-third increase in cortical stromal thickness and a 5-fold increase in subcortical stroma are noted. The increased stroma is due both to hyperplasia of theca cells and to increased formation consequent to the excessive follicular maturation and atresia.
5. There are 4 times more ovarian hilus cell nests (hyperplasia).

Hyperthecosis refers to patches of luteinized theca-like cells scattered throughout the ovarian stroma. It is characterized by the same histologic findings as seen in polycystic ovaries. The clinical picture of the hyperthecosis is a result of overandrogen production. The condition is associated with lower LH levels, which is a possible consequence of the higher testosterone levels blocking estrogen action at the hypothalamic-pituitary level. It seems appropriate to view hyperthecosis as a manifestation of the same process, persistent anovulation, but with greater intensity. A greater degree of insulin resistance is correlated with the degree of hyperthecosis. And in turn, because insulin and IGF-I stimulate proliferation of thecal interstitial cells, hyperinsulinemia may be an important incitement toward hyperthecosis.

The typical histologic changes of the polycystic ovary can be encountered with any size ovary. There is a spectrum of time involved in the development of this condition, and it is useful to view the attainment of large ovaries as a stage of maximal effect of persistent anovulation. Increased size of the ovaries is not a critical feature, nor is it necessary for diagnosis. The key to understanding this clinical problem is an appreciation for the disruption in ovulatory recycling function.

There is no specific pathophysiologic defect. The hypothalamic-pituitary response is entirely appropriate, a response to chronically elevated estrogen feedback. The changes are a functional derangement brought about by accumulated and increased androgen due to a failure of ovulation, whatever the reason. Hence, the polycystic ovary may be associated with extragonadal sources of androgens, or with ovarian androgen-producing tumors, and even with the administration of exogenous androgens to female-to-male transsexuals.

The functional problem can be understood in terms of the two-cell explanation of steroidogenesis (Chapter 6). The follicles are unable to successfully change their microenvironment from androgen dominance to estrogen dominance, the change that is essential for continued follicular growth and development. Measurement of the insulin-like growth factor binding proteins (IGFBPs) in follicular fluid reveals that the profile in polycystic ovaries is the same as that found in atretic follicles, higher levels of IGFBP-2 and -4. This is consistent with limitation of IGF-I and IGF-II activity, reducing the expression of aromatase activity and allowing androgenic dominance in the microenvironment. However, there is a striking difference comparing polycystic granulosa cells with granulosa cells from atretic follicles. Granulosa cells from polycystic ovaries are very sensitive to FSH; granulosa cells from atretic follicles are not. Thus, the granulosa cells from the follicles in polycystic ovaries are not apoptotic (atretic), but simply arrested in development, and capable of responding to FSH stimulation.

The functional picture that emerges (arrested granulosa cells and very active theca cells) corresponds to the morphologic history of underdeveloped granulosa and hyperplastic and luteinized theca. Granulosa cells obtained from the small follicles of polycystic ovaries produce negligible amounts of estradiol but show a dramatic hyperplasia of the theca interstitial cells. In response to the elevated LH levels, the androgen production rate is increased. In turn, because insulin and IGF-I stimulate proliferation of thecal interstitial cells, hyperinsulinemia may be an important incitement toward hyperthecosis.

The familial clustering of anovulation and polycystic ovaries suggests an underlying genetic basis. At least one group of patients with this condition has been described inheriting the disorder, possibly by means of an X-linked dominant transmission. There was a two-fold higher incidence of hirsutism and oligomenorrhea in males when FSH and IGF-I are tested together. In terms of two-cell explanation, this behavior is consistent with a blockage of FSH response (probably through various growth factors), not an intrinsic steroid synthesis enzyme defect. Successful treatment depends, therefore, on altering the ratio of FSH to androgens; either increasing FSH (with clomiphene) or decreasing androgens (wedge resection) to overcome the androgen block at the granulosa level. This permits development of aromatization to bring about conversion of the microenvironment to estrogen dominance. Because anovulation with polycystic ovaries is a functional derangement, it is not surprising that these patients occasionally may ovulate spontaneously. Indeed, ovulation is unpredictable, and contraception may be necessary.

Genetic Considerations

The familial clustering of anovulation and polycystic ovaries suggests an underlying genetic basis. At least one group of patients with this condition has been described inheriting the disorder, possibly by means of an X-linked dominant transmission. There was a two-fold higher incidence of hirsutism and oligomenorrhea in males when FSH and IGF-I are tested together. In terms of two-cell explanation, this behavior is consistent with a blockage of FSH response (probably through various growth factors), not an intrinsic steroid synthesis enzyme defect. Successful treatment depends, therefore, on altering the ratio of FSH to androgens; either increasing FSH (with clomiphene) or decreasing androgens (wedge resection) to overcome the androgen block at the granulosa level. This permits development of aromatization to bring about conversion of the microenvironment to estrogen dominance. Because anovulation with polycystic ovaries is a functional derangement, it is not surprising that these patients occasionally may ovulate spontaneously. Indeed, ovulation is unpredictable, and contraception may be necessary.

A popular hypothesis in the 1990s explained this ovarian hyperandrogenic state as a consequence of an enzymatic dysregulation, specifically of P450c17, the enzyme responsible for both 17a-hydroxylase and 17,20-lyase activities. Abnormal hyperactivity of this enzyme would account for the altered steroidogenesis in both the ovaries and the adrenal glands. Indeed, the adrenal gland is involved in this clinical problem. Higher circulating levels of DHAS, almost exclusively an adrenal product, testify to adrenal participation. The mechanism and the clinical importance of this involvement will be further discussed in Chapter 13. Hirsutism. It is difficult, if not impossible, to know whether abnormal enzyme activity comes first or is a reflection of the anovulatory dysfunctional state. Indeed, in 92 consecutive women with hirsutism, the steroidogenic response to ACTH stimulation was not consistent with an inherent disorder in P450c17. Testing both ovarian and adrenal responses to stimulation fails to delineate a clear-cut or characteristic disorder of P450c17. Genetic screening of anovulatory women with polycystic ovaries has not
The first recognition of an association between glucose intolerance and hyperandrogenism was the famous report of the bearded diabetic woman by Archard and Thiers in 1921. The association between increased insulin resistance and polycystic ovaries is now well recognized. This clinical association of hyperinsulinemia and anovulatory hyperandrogenism is commonly found throughout the world and among different ethnic groups. In addition, hyperinsulinemia and insulin resistance are often associated with acanthosis nigricans. Acanthosis nigricans is a gray—brown velvety, sometimes verrucous, discoloration of the skin, usually at the neck, groin, axillae, and under the breasts, which is a marker for insulin resistance. Hyperkeratosis and papillomatosis are the histologic characteristics of acanthosis nigricans. The presence of acanthosis nigricans in hyperandrogenic women is dependent on the presence and severity of hyperinsulinemia. It is most highly correlated with the magnitude of peripheral insulin resistance and less well with the hyperinsulinemia measured by a glucose tolerance test. The mechanism responsible for the development of acanthosis nigricans is uncertain. Conflicting studies suggest mediation through various growth factor receptors, not just insulin or insulin-like growth factor. Because acanthosis nigricans can be present in normal women, its presence is not an absolute marker for hyperandrogenism and hyperinsulinemia.

Insulin resistance is defined as a reduced glucose response to a given amount of insulin. Resistance to insulin-stimulated glucose uptake is a relatively common phenomenon, sometimes referred to as syndrome X. The majority of patients with noninsulin-dependent diabetes mellitus have peripheral insulin resistance, but not all women who are insulin resistant are hyperandrogenic. The state of chronic hyperinsulinemia represents a compensatory response to the target tissue problem. These relationships involve changes in plasma free fatty acid concentrations. If the insulin levels necessary to suppress free fatty acid levels cannot be achieved, then the increase in free fatty acids leads to increased hepatic glucose production and hyperglycemia.

There are several mechanisms for the state of insulin resistance: peripheral target tissue resistance, decreased hepatic clearance, or increased pancreatic sensitivity. The euglycemic clamp technique establishes a steady state of hyperinsulinemia with a normal glucose level at which point the glucose infusion rate equals glucose utilization. Adding insulin will measure the glucose uptake rate (the more insulin required, the greater the peripheral resistance, also referred to as a measure of insulin sensitivity). Studies with this technique indicate that hyperandrogenic women with hyperinsulinemia have peripheral insulin resistance and, in addition, a reduction in the insulin clearance rate due to decreased hepatic insulin extraction.

The clinical presentation of patients with insulin resistance (whether they have impaired glucose tolerance or diabetes mellitus) depends on the ability of the pancreas to compensate for the target tissue resistance to insulin. At first, compensation is effective, and the only metabolic abnormality is hyperinsulinemia. In many patients, the beta cells of the pancreas eventually fail to meet the challenge, and declining insulin levels lead to impaired glucose tolerance and type II noninsulin-dependent diabetes mellitus. Indeed, beta cell dysfunction can be detected in women with polycystic ovaries even before the appearance of glucose intolerance.

Hyperinsulinemia leads to hypertension and an increased risk of coronary heart disease; a direct relationship exists between plasma insulin levels and blood pressure. Resistance to insulin is further associated with increased triglycerides and decreased HDL-cholesterol levels, a potent bad combination that promotes coronary heart disease. Although the adverse lipid effect is present in anovulatory women with polycystic ovaries, hyperinsulinemia is not encountered in these women until later in life, not in their reproductive years.

Hyperinsulinemia and polycystic ovaries are also associated with an increased production of plasminogen activator inhibitor type 1 (PAI-1). Increased PAI-1 levels are correlated with an increased risk of coronary heart disease; impaired fibrinolysis may affect vascular tissue changes associated with vascular disease.

There are rare causes of hyperinsulinemia and hyperandrogenism that are congenital in origin. Peripheral insulin resistance associated with hyperandrogenism and acanthosis nigricans, known as “type A syndrome,” can be due to mutations in the insulin receptor gene (which leads to decreased numbers of insulin receptors in target tissues). Leprechaunism is a rare syndrome in young girls with a mutation in the insulin receptor gene and perhaps defects in growth factor receptors; it is associated with severe insulin resistance, polycystic ovaries, hyperandrogenism, and acanthosis nigricans. Another subgroup, “type B syndrome,” consists of patients with autoantibodies to insulin receptors. This leaves a large collection of women with neither reduced nor abnormal insulin receptors nor autoantibodies, the most common clinical entity encountered, anovulatory women with hyperandrogenism and hyperinsulinemia. Possible mechanisms for the hyperinsulinemia include functional problems in the insulin receptor (which could also be a consequence of insulin receptor gene mutations) and inhibitors, which can interfere with insulin-receptor function after binding. Thus, there are at least 3 categories for peripheral target tissue insulin resistance: decreased insulin receptor numbers, decreased insulin binding, and post receptor failures.

There is experimental evidence to indicate that in women with polycystic ovaries, the peripheral insulin resistance is due to a defect beyond activation of the receptor kinase, specifically leading to reduced tyrosine autophosphorylation of the insulin receptor. The phosphorylation of serine and threonine residues on the insulin receptor reduces signal transmission, and excessive serine phosphorylation (by a mechanism extrinsic to the insulin receptor) has been demonstrated as a possible postreceptor defect in these patients, changing signal transduction. It has been suggested that serine phosphorylation of the beta chain of the insulin receptor and at the same time of the adrenal and ovarian P450c17 enzyme (the origin or cause of serine phosphorylation is uncertain, but presumably it would have a genetic basis) would explain both the hyperinsulinemia and hyperandrogenism (serine phosphorylation increases and dephosphorylation decreases 17,20-lyase activity and androgen production). Stated simply, serine instead of tyrosine phosphorylation is an “off” mechanism for glucose transport, but an “on” mechanism for P450c17 enzyme activity.
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. Although an initial study reported increased leptin levels in women with polycystic ovaries, the study was criticized for not adjusting for body weight. At least 4 studies, containing for weight, have detected no differences in leptin levels comparing women with and without polycystic ovaries. 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-hour insulin levels in women with polycystic ovaries. Furthermore, a drug that lowers insulin resistance, troglitazone, inhibits transcription of the leptin gene and may be especially suited for obese women with polycystic ovaries.

Is the Link Between Hyperinsulinemia and Hyperandrogenism Explained Solely by the Presence of Obesity in Hyperandrogenic Patients?

Overweight, anovulatory women with hyperandrogenism have a characteristic distribution of body fat known as android obesity. Android obesity is the result of fat deposited in the abdominal wall and visceral mesenteric locations. This fat is more sensitive to catecholamines, less sensitive to insulin, and more active metabolically. This fat distribution is associated with hyperinsulinemia, impaired glucose tolerance, diabetes mellitus, and an increase in androgen production rates resulting in decreased levels of sex hormone-binding globulin and increased levels of free testosterone and estradiol.

Central body (android) obesity is associated with cardiovascular risk factors, including hypertension and unfavorable cholesterol-lipoprotein profiles. The waist:hip ratio is the variable most strongly and inversely associated with the level of HDL, the fraction of HDL-cholesterol most consistently linked with protection from cardiovascular disease. A waist:hip ratio greater than 0.85 indicates android fat distribution. 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. Weight loss in women with lower body obesity is mainly cosmetic, whereas loss of central body weight is more important for general health because an improvement in cardiovascular risk is associated with loss of central body fat.

Hyperinsulinemia and hyperandrogenism, however, are not confined to anovulatory women who are overweight. It is important to note that the combination of increased androgen secretion and insulin resistance has been reported in both obese and nonobese anovulatory women. However, insulin levels are higher and LH, SHBG, and IGFBP-1 levels are lower in obese women with polycystic ovaries compared to nonobese women with polycystic ovaries. For this reason, some have suggested that hyperandrogenic women with polycystic ovaries could be divided into two groups: those with obesity, insulin resistance, hyperinsulinemia, and normal or minimally elevated LH levels; and those with elevated LH, no insulin resistance, and normal insulin levels. In our view, these two groups represent the ends of a spectrum, and division of this clinically broad spectrum of patients is artifactual and unhelpful.

Hyperinsulinemia and hyperandrogenism are not explained, therefore, solely by obesity, and specifically, android obesity. However, the presence of obesity adds the insulin resistance and hyperinsulinemia associated with obesity to that which is specifically unique to the anovulatory, polycystic ovary state.

Which Comes First, the Hyperinsulinemia or the Hyperandrogenism?

There are studies indicating that androgens can induce hyperinsulinemia. However, most of the evidence supports hyperinsulinemia as the primary factor, especially the experiments in which turning off the ovary with a GnRH agonist does not change the hyperinsulinemia or insulin resistance. This indicates that disordered insulin action precedes the increase in androgens. Large doses of insulin were administered to a 16-year-old female with insulin resistance secondary to insulin receptor autoantibodies; the increased insulin levels increased her circulating testosterone levels. With resolution of her insulin resistance, her testosterone levels returned to normal, indicating that the hyperinsulinemia stimulates and increases testosterone and not vice versa. Indeed, there are 6 reasons to believe that hyperinsulinism causes hyperandrogenism:

1. The administration of insulin to women with polycystic ovaries increases circulating androgen levels.
2. The administration of glucose to hyperandrogenic women increases the circulating levels of both insulin and androgens.
3. Weight loss decreases the levels of both insulin and androgens, and increases the levels of IGFBP-1.
4. In vitro, insulin stimulates thecal androgen production.
5. The experimental reduction of insulin levels in women reduces androgen levels in women with polycystic ovaries, but not in normal women.
6. After normalization of androgens with GnRH agonist treatment, the hyperinsulinemia response to glucose tolerance testing remains abnormal in obese women with polycystic ovaries.

Nevertheless, antiandrogen treatment and prolonged androgen suppression can ameliorate the degree of insulin resistance. However, the effect is not great, and may be limited to lean patients with mild hyperinsulinemia.

How Does Hyperinsulinemia Produce Hyperandrogenism?

There is an impressive correlation between the degree of hyperinsulinemia and hyperandrogenism. At higher concentrations, insulin binds to the type I IGF receptors (which are similar in structure to insulin receptors; both IGF and insulin transmit their signals by initiating tyrosine autophosphorylation of their receptors). Thus, when insulin receptors are blocked or deficient in number, it is to be expected that insulin would bind to the type I IGF receptors. In view of the known actions of IGF-I in augmenting the thecal androgen response to LH, activation of IGF-I receptors by insulin would lead to increased androgen production in thecal cells. It should be noted that evidence indicates that the endogenous insulin-like growth factor in the human ovarian follicle is IGF-II in both the granulosa and the thecal cells. 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.
Because the increase in insulin is not always extreme, it has been proposed that insulin activates a signaling system separate from glucose transport, specifically, that insulin operates via inositolphosphoglycan to stimulate steroidogenesis. This pathway would operate by means of insulin binding to its own receptor, not the IGF receptor, a pathway supported by in vitro studies of both granulosa cells and thecal cells.

There are two other important actions of insulin which contribute to hyperandrogenism in the presence of hyperinsulinemia: inhibition of hepatic production of insulin-like growth factor binding protein-1.

Independently of any effect on sex steroids, increased insulin will inhibit the hepatic synthesis of sex hormone-binding globulin. In vitro studies indicate that both insulin and IGF-I directly inhibit SHBG secretion by human hepatoma cells. This is now known to be the mechanism for the inverse relationship between body weight and the circulating levels of SHBG. Because SHBG is regulated by insulin, decreased SHBG levels in women represent an independent risk factor for noninsulin-dependent diabetes mellitus, regardless of body weight and fat distribution. Of course, a decrease in SHBG allows more androgen and estrogen to be bioavailable.

Nutritional intake decreases the circulating levels of insulin-like growth factor binding protein-1 (IGFBP-1) because of the increase in insulin, which then directly inhibits IGFBP-1 production in the liver. Obese individuals with increased insulin levels and women with polycystic ovaries and hyperinsulinemia have lower circulating levels of IGFBP-1. This lower level of IGFBP-1 allows an increase in circulating levels of IGF-I and greater local activity of IGF-I and/or IGF-II in the ovary. In addition, greater IGF-I activity in the endometrium due to reduced levels of IGFBP-1 and direct insulin activation of IGF receptors or its own receptor are possible mechanisms for endometrial growth (and the increased risk for endometrial cancer) in these patients. On the other hand, a greater IGF binding protein capacity in the follicular fluid from polycystic ovaries would yield a reduced bioavailability of IGFs within the follicle. Although these findings (increased circulating IGF availability and decreased follicular IGF availability) at first seem paradoxical, they are compatible with increased thecal androgen production by the IGF pathway and disrupted follicular maturation by the FSH system. It is likely that these characteristics of polycystic ovaries are secondary to increased anovulation, hyperinsulinemia, and increased androgens, rather than indicating a primary, etiologic role.

Finally, there is evidence that insulin can increase LH secretion in some anovulatory, overweight women.

Why are not all patients with hyperinsulinemia also hyperandrogenic; e.g., many women with noninsulin-dependent diabetes mellitus? The answer to this question is not known, but a logical speculation is that an ovarian genetic susceptibility is required, although it may be that the existence of long-term anovulation must be present and even precede hyperinsulinemia.

Can Ovarian Hyperandrogenism be Treated with Diabetes Drugs?

There is a variety of pharmacologic agents available to reduce insulin levels. Diazoxide and octreotide, the long-acting analogue of somatostatin, both inhibit insulin secretion, but are accompanied by worsening glucose intolerance. Metformin and troglitazone, oral agents used to treat diabetes mellitus, have been administered to anovulatory women with polycystic ovaries.

Metformin improves insulin sensitivity, but the primary effect is a significant reduction in gluconeogenesis, thus decreasing hepatic glucose production. Metformin treatment (500 mg tid) reduces hyperinsulinemia, basal and stimulated LH levels, free testosterone concentrations, and PAI-1 levels in overweight women with polycystic ovaries. A significant number of these anovulatory women ovulate and achieve pregnancy. In a group of obese women with polycystic ovaries, 90% of the women treated with metformin and 50 mg clomiphene ovulated compared with 8% in the group treated with placebo and clomiphene. However, there has been controversy, suggesting that the improvement was the result of the weight loss that often accompanies the use of metformin. In a study designed to control the effect of body weight, the administration of metformin was without effect on insulin resistance in extremely overweight women with polycystic ovaries. In another well-designed study, metformin again had no effect on insulin resistance when body weights remained unchanged, and in this study baseline weights and hyperinsulinemia were only modestly increased. In lean, anovulatory women with hyperinsulinemia, metformin treatment reduced hyperandrogenemia although there was no change in body weight; however, a decrease in the waist to hip ratio accompanied a reduction in the hyperinsulinemia. This study indicates that both obese and nonobese patients with hyperinsulinemia respond to metformin treatment. The reasons for the differences among the studies are not apparent. Perhaps
An increased risk of endometrial cancer and, perhaps, breast cancer. Hirsutism, alopecia, and acne

The lipid 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

Besides contraception, there is another argument in favor of continuous suppression with low-dose oral contraceptives rather than periodic progestational interruption.

Anovulation is the key feature of this condition and presents as amenorrhea in approximately 50% of cases and with irregular, heavy bleeding (dysfunctional uterine bleeding) in 30%. True virilization is rare, but 70% of anovulatory patients complain of cosmetically disturbing hirsutism. The development of hirsutism depends not only on the concentration of androgens in the blood but on the genetic sensitivity of hair follicles to androgens. Thus, anovulatory, hyperandrogenic women can be free of the clinical sign of hirsutism. Aloppecia and acne can also be consequences of hyperandrogenism. Obesity has been classically regarded as an important feature, but in view of the concept of persistent anovulation arising from many causes, its presence is extremely variable (about 35–60% of anovulatory women with polycystic ovaries) and has no diagnostic value. However, the greater the body mass index, the higher the testosterone levels, and, therefore, hirsutism is more common in overweight anovulatory women.

Although an elevated LH value in the presence of a low or low-normal FSH may be diagnostic, the diagnosis is easily made by the clinical presentation alone. About 20–40% of patients with this condition do not have elevated LH levels with reversal of the LH:FSH ratio. We do not routinely measure FSH and LH levels in anovulatory patients. The selection of a specific criterion to make this diagnosis will inevitably fail to include many patients with this clinical problem that covers a broad spectrum of manifestations. This failing also applies to the diagnostic use of ultrasonographic criteria.

The Clinical Consequences of Persistent Anovulation


There are potentially severe clinical consequences of the steady state of hormone secretion. Besides the problems of bleeding, amenorrhea, hirsutism, and infertility, the effect of the unopposed and uninterrupted estrogen is to place the patient at considerable risk for cancer of the endometrium and, perhaps, cancer of the breast.

The risk of endometrial cancer is increased 3-fold, while chronic anovulation during the reproductive years has been reported to be associated with a 3–4 times increased risk of breast cancer appearing in the postmenopausal years. However, the statistical power of these observational studies on breast cancer was limited by small numbers (fewer than 15 cases). Others have failed to find a link between anovulation and the risk of breast cancer.

If left unattended, patients with persistent anovulation develop clinical problems, and, therefore, appropriate therapeutic management is essential for all anovulatory patients. In a long-term follow-up of women with polycystic ovaries, the problems of android obesity and hyperinsulinemia were observed to persist into the postmenopausal years. Although it is yet to be documented by appropriate epidemiologic studies, it is logical to expect postmenopausal women who have previously been anovulatory, hyperandrogenic, and hyperinsulinemic to experience a reduction in life expectancy because of cardiovascular disease and diabetes mellitus. These women will derive important benefits from an aggressive preventive health care attitude on the part of the clinician that results in amelioration of adverse metabolic risk factors.

Growing support for an inherited basis for this disorder (autosomal dominant) makes it important to consider appropriate family counseling. Sisters and daughters may have a 50% chance of having the same problems as the patient.

The typical patient presents with anovulation and irregular menses or amenorrhea with withdrawal bleeding after a postgestational challenge. Documentation of anovulation is usually unnecessary, especially in the presence of menstrual irregularity with periods of amenorrhea. In the patient who has long-standing anovulation, an idiopathic biopsy (with extensive sampling) is a wise precaution. The well-known association between this condition and abnormal endometrial changes must be kept in mind. Endometrial cancer can be encountered in young, anovulatory women.

The decision to perform an endometrial biopsy should not be influenced by the patient's age. It is the duration of exposure to unopposed estrogen that is critical. Therapy of any anovulatory patients can be planned at the first visit. If the patient desires pregnancy, she is a candidate for the medical induction of ovulation. Therapy of any anovulatory patients can be planned at the first visit. If the patient desires pregnancy, she is a candidate for the medical induction of ovulation. Therapy of any anovulatory patients can be planned at the first visit. If the patient desires pregnancy, she is a candidate for the medical induction of ovulation (Chapter 30). When pregnancy is achieved, patients with polycystic ovaries appear to have an increased risk of spontaneous miscarriage. This increased risk has been attributed to elevated levels of LH that may produce an adverse environment for the oocyte, perhaps even inducing premature maturation and completion of the first meiotic division. For this reason, consideration should be given to pretreatment suppression prior to the induction of ovulation (Chapter 30). If the patient presents with amenorrhea, an investigation must be pursued as outlined in Chapter 11. The management of significant dysfunctional uterine bleeding is discussed in Chapter 16 and hirsutism in Chapter 13.

For the patient who does not wish to become pregnant and does not complain of hirsutism, but is anovulatory and has irregular bleeding, therapy is directed toward interruption of the steady state effect on the endometrium and breast. The use of medroxyprogesterone acetate (5–10 mg daily for the first 10 days of every month) is favored to ensure complete withdrawal bleeding and to prevent endometrial hyperplasia and atypia. The monthly 10-day duration has been demonstrated to be essential to protect the endometrium from cancer in women on postmenopausal estrogen therapy. Until specific clinical data are available, it seems logical that young, anovulatory women also require at least 10 days of postovulatory exposure every month. The patient will be aware of the onset of ovulatory cycles because bleeding will occur at a time other than the expected withdrawal bleed. In our opinion, when reliable contraception is essential, the use of low-dose combination oral contraception in the usual cyclic fashion is appropriate.
Although the elevated androgens associated with polycystic ovaries and anovulation offer some protection against osteoporosis, the adverse impact on insulin resistance at an increased rate of resumption of ovulation and pregnancy. Hyperinsulinemic women at greater risk for noninsulin-dependent diabetes, but the age of onset is about 30 years earlier than the general population. A contributing factor to the abnormal lipid pattern in many of these patients is hyperinsulinemia. A goal for weight loss that correlates with a good chance of achieving pregnancy (improved menstrual function), a reduction in insulin levels, and a decrease in free testosterone levels is a body mass index less than 27 (see the nomogram for calculating body mass index in Chapter 19). In a long-term follow-up study, anovulatory women with polycystic ovaries had a 5-fold increased risk of diabetes mellitus compared with age-matched controls. It is appropriate, and indeed essential, to monitor glucose tolerance with periodic glucose tolerance testing.

Hyperinsulinemia also contributes to the increased risk of cardiovascular disease both by means of a direct atherogenic action and indirectly by adversely affecting the lipoprotein profile. Insulin resistance may be a more significant factor than androgens in determining the abnormal lipoprotein profile in overweight, anovulatory women. It has also been suggested that the increased insulin stimulation of IGF-I could produce bone changes similar to that seen in acromegaly. Hyperinsulinemia may be a factor contributing to the higher risk of endometrial cancer in these patients by increasing IGF-I activity in the endometrium. The best therapy for these women is weight loss. Both the hyperinsulinemia and the hyperandrogenism can be reduced with weight loss, which is at least more than 5% of the initial weight. Increased PAI-1 levels associated with hyperinsulinemia also improve with weight loss. These metabolic improvements are associated with an impressive rate of resumption of ovulation and pregnancy. In one study, 60 of 67 anovulatory women, who lost from 4 to 15 kg, resumed ovulation. A goal for weight loss that correlates with a good chance of achieving pregnancy (improved menstrual function), a reduction in insulin levels, and a decrease in free testosterone levels is a body mass index of less than 27 (see the nomogram for calculating body mass index in Chapter 19). In a long-term follow-up study, anovulatory women with polycystic ovaries had a 5-fold increased risk of diabetes mellitus compared with age-matched controls. It is appropriate, and indeed essential, to monitor glucose tolerance with periodic glucose tolerance testing.

Hyperinsulinemia also contributes to the increased risk of cardiovascular disease both by means of a direct atherogenic action and indirectly by adversely affecting the lipoprotein profile. Insulin resistance may be a more significant factor than androgens in determining the abnormal lipoprotein profile in overweight, anovulatory women. It has also been suggested that the increased insulin stimulation of IGF-I could produce bone changes similar to that seen in acromegaly. Hyperinsulinemia may be a factor contributing to the higher risk of endometrial cancer in these patients by increasing IGF-I activity in the endometrium. Hyperinsulinemia may be a factor contributing to the higher risk of endometrial cancer in these patients by increasing IGF-I activity in the endometrium.
Overall Goals of Treatment
1. Reduce the production and circulating levels of androgens.
2. Protect the endometrium against the effects of unopposed estrogen.
4. Lower the risk for cardiovascular disease.
5. Avoid the effects of hyperinsulinemia on the risks of cardiovascular disease and diabetes mellitus.
6. Induction of ovulation to achieve pregnancy.

Who Should Be Tested for Hyperinsulinemia?

Do all anovulatory patients require testing for hyperinsulinemia? Both lean and obese women with polycystic ovaries can be found to have hyperinsulinemia, but not all anovulatory women with polycystic ovaries (lean and obese) have hyperinsulinemia. However, it is more common and severe in overweight women and androgenic effects are more intense. Lean women with hyperinsulinemia do not appear to have the same risk of future diabetes mellitus, although clinical follow-up may in time document an onset later in life of noninsulin-dependent diabetes mellitus compared to an earlier onset in obese women. It has been reported that anovulatory women with polycystic ovaries who have a parent with noninsulin-dependent diabetes mellitus are more likely to have impaired beta cell secretion of insulin in addition to insulin resistance.

Because of the probable inherited susceptibility for anovulation and insulin resistance, consideration should be given to a glucose tolerance and insulin evaluation for family members of already diagnosed patients. Both brothers and sisters of anovulatory, hyperandrogenic women can be insulin resistant.

What about those women who are ovulatory and have no clinical complaints, yet supposedly have an underlying hyperinsulinemic disorder? In our view, if this is real, it is a homeostatic compensatory state, and until appropriate data reveal adverse outcomes in these women, diagnostic and therapeutic interventions are not indicated.

It would be ideal if all patients with android obesity were tested for hyperinsulinemia. The waist:hip ratio is a means of estimating the degree of upper to lower body obesity; the ratio accurately predicts the amount of intra-abdominal fat (which is greater with android obesity). However, studies have demonstrated that the more easily determined circumference of the waist is a better predictor of central, android abdominal fat. A waist circumference greater than 100 cm (about 35 inches) in men and 90 cm (about 36 inches) in women is predictive of abnormal endocrinologic and metabolic function and is associated with an increased risk of cardiovascular disease.

Teenagers who present with persistent anovulation would also be good candidates for hyperinsulinemia testing. During puberty, insulin resistance develops, probably because of the increase in sex steroids and growth hormone, resulting in a secondary increase in insulin and IGF-I. The increase in insulin leads to a decrease in SHBG, which would allow greater sex steroid activity for pubertal development. There is reason to believe that some teenagers fail to normalize the hyperinsulinemia associated with the growth hormone increase in early puberty. It would be important to identify these teenagers who are at an increased risk for the development of diabetes mellitus and are destined to struggle with all of the problems associated with anovulation and polycystic ovaries. There is also reason to believe that many cases of premature adenarche are due to hyperinsulinemia, and these patients go on to develop the full characteristics of anovulation, hyperandrogenism, and polycystic ovaries.

When clinical circumstances are encountered that do not seem to make sense, give consideration to the presence of hyperinsulinemia. We have experienced several examples where hyperinsulinemia proved to be the underlying answer. Several instances involved women in the reproductive age range, with significant hirsutism and very high testosterone levels. Evaluation failed to reveal an ovarian tumor or an adrenal lesion. The demonstration of hyperinsulinemia has avoided unnecessary surgical exploration for a “small tumor.” Another example was the onset of hirsutism in elderly women associated with testosterone levels greater than 200 ng/mL and normal imaging evaluations of the ovaries and adrenal glands. Again, hyperinsulinemia was the cause, not a hidden tumor.

Unfortunately, it is not certain what levels of insulin in the fasting state or in response to an oral glucose tolerance test are correlated with clinical outcome. However, in individuals with normal glucose tolerance, the fasting insulin level is strongly correlated with insulin resistance. In most laboratories, the upper limit of normal for a fasting insulin level is 10–20 U/mL. Because there is considerable overlap between normal women and patients with anovulation and polycystic ovaries, it is reasonable to assume that all overweight, anovulatory women with polycystic ovaries are hyperinsulinemic. Nevertheless we recommend the measurement of the ratio of fasting glucose to fasting insulin in order to provide evidence that lends credence and importance to counseling efforts. A ratio of less than 4.5 is consistent with insulin resistance.

Based upon the information reviewed in this chapter, we offer the following recommendations:

All anovulatory women who are hyperandrogenic should be assessed for insulin resistance and glucose tolerance with measurements of:

1. The fasting glucose:insulin ratio, followed by
2. The 2-hour glucose level after a 75 g glucose load:
   - Normal: less than 140 mg/dL
   - Impaired: 140–199 mg/dL
   - Noninsulin-dependent diabetes mellitus: 200 mg/dL or higher

Anovulatory women who do not exhibit signs of hyperandrogenism should be evaluated for the presence of a metabolic abnormality by measuring the free testosterone level; if elevated, insulin resistance and glucose tolerance should be assessed. However, a fasting glucose:insulin ratio is about one-third the cost of a free testosterone level. It may be more economical to measure this ratio in all anovulatory women.

In women who continue to manifest this disorder, periodic surveillance is necessary. The frequency is uncertain, but annual assessment is appropriate in women who continue to be obese.

Conclusion

We are truly entering a new era in our understanding and management of women with polycystic ovaries and hyperandrogenism. In the past we have effectively treated the specific problems of infertility, dysfunctional uterine bleeding, and hirsutism. We now have an opportunity to have an impact on the quality and quantity of life to be experienced by these patients. By creating and supporting a preventive health care attitude in anovulatory women, we can not only correct specific clinical consequences of anovulation, we can reduce major adverse effects on overall health.

Chapter References

in women with polycystic ovary syndrome. Metabolism 44:525, 1995.


246. Lagro RS, Finegood D, Dunaff A, A fasting glucose to insulin ratio is a useful measure of insulin sensitivity in women with polycystic ovary syndrome, J Clin Endocrinol Metab 83:2694, 1998.


248. Govind A, Obhral MS, Clayton RN, Polycystic ovaries are inherited as an autosomal dominant trait: analysis of 29 polycystic ovary syndrome and 10 control families, J Clin Endocrinol Metab 84:38, 1999.

Copyright ©
Hirsutism, excessive facial and body hair caused by excess androgen production, is usually associated with anovulatory ovaries and loss of cyclic menstrual function. The more severe states of virilism (clitoromegaly, deepening of the voice, balding, and changes in body habitus) are rarely seen and usually are secondary to adrenal hyperplasia or androgen-producing tumors of adrenal or ovarian origin. Although these are rare, diagnostic evaluation is required.

A concerned and sympathetic approach must be offered to women who complain of hirsutism. The responsible clinician must view hirsutism both as an endocrine problem and as a cosmetic problem. To the affected woman, hair growth over the face, abdomen, or breasts is disturbing on several levels. Is there disease? Is sexuality changing? Is social acceptance altered? Is fertility impaired?

This chapter reviews the biology of hair growth and the endocrine causes that can yield hirsutism. An uncomplicated, effective program for diagnostic evaluation and therapeutic management is presented.

The Biology of Hair Growth

Embryology

Each hair follicle develops at about 8–10 weeks of gestation as a derivative of the dermis. It is composed initially of a solid column of cells that proliferates from the basal layers of the epidermis and protrudes downward into the dermis. As the column elongates it encounters a cluster of mesodermal cells (the dermal papilla) that it envelops at its bulbous tip (bulb). The solid epithelial column then hollows out to form a hair canal, and the pilosebaceous apparatus (a hair follicle, sebaceous glands, and arrector pili muscles) is laid down.

![Diagram of a hair follicle](image)

Hair growth begins with proliferation of the epithelial cells at the base of the column in contact with the dermal papilla. The lanugo hair that covers the fetus is lightly pigmented, thin in diameter, short in length, and fragile in attachment. Important to note is the fact that the total endowment of hair follicles is made at an early gestational stage (by 22 weeks) and no new hair follicles will be produced de novo. The concentration of hair follicles laid down per unit area of facial skin does not differ materially between sexes but does differ between races and ethnic groups (Caucasian > Oriental, Mediterranean > Nordic). In addition, hair growth differences between races probably reflect hair follicle differences in 5a-reductase activity (the production of the active androgen, dihydrotestosterone).

The pattern of hair growth is genetically predetermined.

Structure and Growth

Hair does not grow continuously but, rather, in a cyclic fashion with alternating phases of activity and inactivity. The cycles are referred to by the following terms:

- **Anagen** — the growing phase.
- **Catagen** — rapid involution phase.
- **Telogen** — quiescent phase.
In the resting phase (telogen), the hair is short and loosely attached to the base (the bulb) of the epithelial canal. As growth begins (anagen), epithelial matrix cells at the base begin to proliferate and extend downward into the dermis. The epithelial column elongates some 4–6 times from the resting state. Once downward extension is completed, continued rapid growth of the matrix cells pushes upward to the skin surface. The tenuous contact of the previous hair is broken, and that hair is shed. The superficial matrix cells differentiate forming a keratinized column. Growth continues as long as active mitoses persist in the basal matrix cells. When finished (catagen), the column shrinks, the bulb shrivels, and the resting state is reached (telogen).

The length of hair is primarily determined by duration of the growth phase (anagen). Scalp hair remains in anagen for 3 years and has only a relatively short resting phase. Elsewhere (forearm) a short anagen and long telogen will lead to short hair of stable nongrowing length. The appearance of continuous growth (or periodic shedding) is determined by the degree to which individual hair follicles act asynchronously with their neighbors. Scalp hair is asynchronous and, therefore, always seems to be growing. The resting phase that some hairs (10–15%) are in is not apparent. If marked synchrony is achieved, then all hairs may undergo telogen at the same time leading to shedding that is called **telogen effluvium**. Occasionally, women will complain of marked hair loss from the scalp, but this time period of shedding is usually limited (6–8 months), and growth resumes when asynchrony is re-established. **However, it is worthwhile to rule out thyroid disease with a TSH measurement.**

**Hypertrichosis**: a generalized increase in hair of the fetal lanugo type, associated with the use of drugs or malignancy. **Vellus hair** is the downy, unpigmented hair associated with the pubertal years. **Terminal hair** is the coarse, pigmented hair that grows on various parts of the body during the adult years. Hirsutism implies a vellus to terminal hair transformation.

**Factors That Influence Hair Growth**

The dermal papilla is the director of the events that control hair growth. Despite major injury to the epithelial component of the follicle (such as freezing, x-rays, or a skin graft), if the dermal papilla survives, the hair follicle will regenerate and regrow hair. Injury to or degeneration of the dermal papilla (such as that achieved with properly performed electrolysis) is the crucial factor in permanent hair loss.

Sexual hair is defined as hair that responds to the sex steroids. Sexual hair grows on the face, lower abdomen, anterior thighs, the chest, the breasts, the pubic area, and in the axillae. Once androgen influences hair follicles in sexual areas and larger, longer, more pigmented hair is induced, these final hair characteristics recur in typical cycles of activity and inactivity, even in the absence of sustaining androgen. Androgenic stimulation of the hair follicle requires the conversion of testosterone to dihydrotestosterone, and therefore the sensitivity of the hair follicle to androgens is determined by the level of 5a-reductase activity. The individual variability in hair response is believed to reflect genetic differences in 5a-reductase activity.

From animal studies and human disease patterns, the following list of hormonal effects can be compiled:

1. Androgens, particularly testosterone, initiate growth, increase the diameter and pigmentation of the keratin column, and probably increase the rate of matrix cell mitoses in all but scalp hair.
2. Estrogens act essentially opposite from androgens, retarding the rate and initiation of growth, and leading to finer, less pigmented and slower growing hair. In pregnancy (high estrogen and progesterone) can increase the synchrony of hair growth, leading to periods of growth or shedding.
3. Progestins have minimal direct effects on hair.
4. Pregnancy (high estrogen and progesterone) can increase the synchrony of hair growth, leading to periods of growth or shedding.

An important clinical characteristic of hair growth can be understood from studies of the effects of castration. If castration occurs before puberty, the male will not grow a beard. If castration occurs after puberty with beard and sexual hair distribution fully developed, then these hairs continue to grow albeit more slowly and with finer caliber. Androgen stimulates sexual hair follicle conversion from lanugo to terminal adult hair growth patterns, but **once established, these patterns persist despite withdrawal of androgen.**

Sexual and nonsexual hair growth can be affected by endocrine problems. In hypopituitarism, there is marked reduction of hair growth. Acromegaly will be associated with hirsutism in 10–15% of patients. Although the impact of thyroid hormone is not clear, hypothyroid individuals sometimes display scalp alopecia as well as less axillary, pubic, and curiously, lateral eyebrow hair. 5a-Reductase activity is stimulated by insulin-like growth factor-I (IGF-I). Increased IGF-I activity in anovulatory patients with insulin resistance and hyperinsulinemia can intensify the hirsute response in these hyperandrogenic patients.

Hair growth can be influenced by nonhormonal factors, such as local skin temperature, blood flow, and edema. Hair grows faster in the summer than in the winter. Hair growth can be seen in association with central nervous system (CNS) pathology such as encephalitis, cranial trauma, multiple sclerosis, and with certain drugs.

**Androgen Production**

The production rate of testosterone in the normal female is 0.2 to 0.3 mg/day. Approximately 50% of testosterone is derived from peripheral conversion of androstenedione, whereas the adrenal gland and ovary contribute approximately equal amounts (25%) to the circulating levels of testosterone, except at midcycle when the ovarian contribution increases by 10–15%. Dehydroepiandrosterone sulfate (DHAS) arises almost exclusively from the adrenal gland, while 90% of dehydroepiandrosterone (DHA) is from the adrenal.
About 80% of circulating testosterone is bound to a beta-globulin known as sex steroid hormone-binding globulin (SHBG). In women, approximately 19% is loosely bound to albumin, leaving about 1% unbound. Androgenicity is dependent mainly on the unbound fraction and partly upon the fraction associated with albumin. DHA, DHAS, and androstenedione are not significantly protein bound, and routine immunoassay reflects their biologically available hormone activity. This is not the case with testosterone because routine assays measure the total testosterone concentration, bound and unbound.

SHBG production in the liver is decreased by androgens. Hence, the binding capacity in men is lower than in normal women, and 2–3% of testosterone circulates in the free, active form in men. SHBG is decreased by insulin, and increased by estrogens and thyroid hormone. Therefore, binding capacity is increased in women with hyperthyroidism, in pregnancy, and by estrogen-containing medication. In a hirsute woman, the SHBG level is depressed by the excess androgen (and, when present, by hyperinsulinemia), and the percent free and active testosterone is elevated as is the metabolic clearance rate of testosterone. The total testosterone concentration, therefore, can be in the normal range in a woman who is hirsute. However, there is little clinical need for a specific assay for the free portion of testosterone. The very presence of hirsutism or masculinization indicates increased androgen effects. One can reliably interpret a normal total testosterone level in these circumstances as being compatible with decreased binding capacity and increased free testosterone.

In hirsute women, only 25% of the circulating testosterone arises from peripheral conversion, and most is due to direct glandular secretion. Indeed, data overwhelmingly indicate that the ovary is the major source of increased testosterone and androstenedione in hirsute women. The most common cause of hirsutism in women is persistent anovulation and excessive androgen production by the ovaries. Adrenal causes are most uncommon.

3α-Androstanediol Glucuronide

Although testosterone is the major circulating androgen, dihydrotestosterone (DHT) is the major nuclear androgen in many sensitive tissues, including hair follicles and the piceosebaceous unit in skin. 3α-Androstanediol is the peripheral tissue metabolite of DHT, and its glucuronide, 3α-androstanediol glucuronide (3α-AG) has been utilized as a marker of target tissue cellular action. There is an excellent correlation between the serum levels of 3α-AG and the clinical manifestations of androgens. Specifically, 3α-AG correlates with the level of 5α-reductase activity (testosterone and androstenedione to dihydrotestosterone) in the skin.

Thus, there are 3 principal laboratory measurements of potential clinical use for the evaluation of androgen excess:

1. Testosterone — a measure of ovarian and adrenal activity.
2. DHAS — a measure of adrenal activity.
3. 3α-AG — a measure of peripheral target tissue activity.

Hirsutism is not a disorder of hair; rather, it reflects increased 5α-reductase activity which produces more DHT, leading to the stimulation of hair growth. This enzyme activity is increased by an increased availability of precursor (the circulating level of testosterone, therefore, is a primary factor) or by still unknown local tissue mechanisms. Measurement of 3α-AG has revealed that true idiopathic hirsutism may not exist (or at least it is very rare). In the presence of other laboratory measurements that are normal, increased levels of 3α-AG indicate an increased activity of 5α-reductase in the peripheral compartment. However, 3α-AG also reflects hepatic conjugation activity and the impact of major precursors that are derived from the adrenal gland and not from peripheral sources. Therefore 3α-AG is not solely a measure of cutaneous androgen metabolism.

There are 2 reasons why the measurement of 3α-AG is not part of the routine clinical approach to the problem of hirsutism. First, it is not an absolute measurement. Values in hirsute women overlap the normal range by about 20%. Second, and most importantly, the ultimate diagnosis and therapy of the problem are not affected by this test.

Prostate-Specific Antigen

The prostate-specific antigen (PSA) is a serine protease, produced in the prostate gland, and used as a tumor marker for the diagnosis and management of prostate cancer. PSA has also been detected in female tissues, and using a very sensitive assay, circulating levels of PSA can be measured in women. Because androgens increase the expression of the PSA gene, it is reasonable to expect hyperandrogenic women to have increased circulating levels of PSA. Indeed, circulating levels of PSA are higher in hirsute women and are correlated with 3α-AG levels. These elevated levels do not respond to ovarian and adrenal stimulation or ovarian suppression, indicating a widespread origin. However, at the present time, there is no clinical application for PSA measurements in women.

Evaluation of Hirsutism

Cosmetically disturbing hirsutism is the end result of a number of factors:

1. The number of hair follicles present (Asian women bearing androgen-producing tumors rarely are hirsute because of the low concentration of hair follicles per unit skin area).
2. The degree to which androgen has converted resting vellus hair to terminal adult hair.
3. The ratio of the growth to resting phases in affected hair follicles.
4. The asynchrony of growth cycles in aggregates of hair follicles.
5. The thickness and degree of pigmentation of individual hairs.

The primary factor in hirsutism is an increase in androgen levels (usually testosterone) that produces an initial growth stimulus and then acts to sustain continued growth. Essentially, every woman with hirsutism will have an increased production rate of testosterone and androstenedione.

Anovulatory women who are not hirsute can often be found to have laboratory evidence of increased androgen production. However, the most common complaint voiced by women and associated with increased androgen production is hirsutism. This is followed in order by acne and increased oiliness of the skin, increased libido, clitoromegaly, and, finally, masculinization. Masculinization and virilization are terms reserved for extreme androgen effects (usually, but not always, associated with a tumor) leading to the development of a male hair pattern, clitoromegaly, deepening of the voice, increased muscle mass, and general male-like body habitus. The Ferriman-Gallwey scoring system used to quantify the amount of hirsutism is of little clinical use, and it is reserved for studies of hirsutism, but even for this...
purpose, it is limited by subjective variability.  

Alopecia can be a vexing problem for patient and clinician. In many instances, alopecia is a temporary phenomenon, a response of the scalp hair to some change that has induced a period of synchronous hair growth and loss. This can be a response to acute, stressful events. Telogen effluvium often occurs near the end of a pregnancy or postpartum. With time, usually 6 months to a year, the scalp hair becomes asynchronous again and the hair thins. In a series of consecutive patients presenting with diffuse alopecia, the majority had no evidence of hirsutism or menstrual dysfunction; however, anovulation with polycystic ovaries was the most common problem, and nearly 40% demonstrated hyperandrogenism. Patients complaining of alopecia deserve evaluation for hyperandrogenism because a significant number of them can be appropriately treated. In addition, a laboratory survey is indicated for thyroid dysfunction or chronic illness. However, because alopecia reflects increased scalp 5α-reductase activity, normal circulating hormone levels should not preclude treatment. Hair loss is also a consequence of aging, beginning in both sexes about the age of 50.

Acne is another sign of increased androgen activity. Up to 60% of women with acne who have normal circulating levels of androgens display evidence of increased 5α-reductase activity in the pilosebaceous unit. These are the women who benefit from antiandrogenic treatment.

Acanthosis nigricans in an overweight patient with hirsutism is a reliable clinical marker of insulin resistance and hyperinsulinemia. This gray—brown velvety discoloration of the skin is usually present at the neck, groin, and axillae; however, the vulva is a very common site in hirsute women. Acanthosis nigricans indicates the need to determine the state of glucose metabolism, as discussed in Chapter 12. Serious consideration should be given to the presence of hyperinsulinemia in hyperandrogenic women.

The most common clinical problem is the hirsute woman with irregular menses, with the onset of hirsutism during teenage years or in the early 20s, and long, gradual worsening of the condition. About 70% of anovulatory women develop hirsutism. The picture is so characteristic that a careful history is often sufficient for the diagnosis.

A good history can reveal some of the rare causes of hirsutism: environmental factors producing chronic irritation or reactive hyperemia of the skin, the use of drugs, changes associated with Cushing's syndrome or acromegaly, or even the presence of pregnancy (indicating the possibility of a luteoma). Hair-stimulating drugs include methyltestosterone, anabolic agents like Nilevar or Anavar, phenytoin, diazoxide, danazol, cyclosporin, and minoxidil. Hirsutism associated with drugs that are not androgens typically consists of fine hairs distributed diffusely over the trunk and face (hypertrichosis). The 19-nortestosterones in the current low-dosage oral contraceptives rarely (if ever) cause acne or hirsutism. Postmenopausal women being treated with androgens can develop hirsutism despite low doses. Dehydroepiandrosterone, available as a food supplement, increases testosterone levels in women and can cause hirsutism and acne, even at the lowest doses.

Especially important in the history is the rapidity of development. A woman who develops hirsutism after the age of 25 and demonstrates very rapid progression of masculinization over several months usually has an androgen-producing tumor.

Late-onset adrenal hyperplasia caused by an enzymatic deficiency presenting in adult life is rare. Classical congenital adrenal hyperplasia that can lead to hirsutism is usually diagnosed and treated prior to puberty. Hirsutism in childhood is usually caused by classical congenital adrenal hyperplasia or androgen-producing tumors. Genetic problems, such as Y-containing mosaics or incomplete androgen sensitivity, will produce signs of androgen stimulation at puberty.

Vilirization during pregnancy raises the suspicion of a luteoma, which is not a true tumor but an exaggerated reaction of the ovarian stroma to normal levels of chorionic gonadotropin. The solid luteoma is unilateral in 45% of cases and associated with a normal pregnancy. Theca-lutein cysts (also called hyperreactio luteinalis) seen with trophoblastic disease are virtually always bilateral. Maternal vilirization occurs in 30% of pregnancies with theca-lutein cysts. Hyperreactio luteinalis can also be seen with the high human chorionic gonadotropin (HCG) liters associated with multiple gestation. Because a luteoma regresses postpartum, the only risk is masculinization of a female fetus; a risk not reported with theca-lutein cysts. Luteomas cause maternal vilirization in 35% of the cases, and in these pregnancies, about 80% of female fetuses will exhibit some signs of masculinization. Subsequent pregnancies are usually normal, but maternal vilirization is occasionally recurrent. There have been several rare case reports of recurrent maternal vilirization during pregnancy associated with hyperthyecosis or polycystic ovaries. These cases should also be regarded as examples of hyperreactio luteinalis.

Androgen-secreting ovarian tumors are very rarely encountered during pregnancy, probably because the excess androgen usually suppresses ovulation. Ultrastructural evaluation of the pelvis in women experiencing vilirization in pregnancy is very helpful. Malignancy is frequently encountered when a solid unilateral ovarian lesion is present.

Hirsutism, therefore, is usually associated with persistent anovulation (Chapter 12). Although anovulatory ovaries are usually the source of excess androgens, a minimal workup is necessary, dedicated to ruling out the adrenal sources and tumors. It should be emphasized that hospitalization for extensive evaluation of hirsutism is required only rarely.

The Diagnostic Workup for Hirsutism

The initial laboratory evaluation of hirsutism consists of assays for the blood levels of testosterone and 17α-hydroxyprogesterone (17-OHP). We no longer consider the measurement of DHAS to be necessary (discussed later). As part of the evaluation for anovulation, prolactin levels and thyroid function should be assessed, careful examination of the breasts for the presence of galactorrhea is important, and an aspiration endometrial biopsy should be considered. A thyroid-stimulating hormone (TSH) screen is also indicated in women complaining of alopecia. In addition, screen is also should be directed to the possible presence of hyperinsulinemia. Patients with intense androgen action may be amenorrheic due to endometrial suppression (with a decidual response) and may not demonstrate withdrawal bleeding after a progestational challenge.

Cushing's syndrome can present with hirsutism, and later, masculinization. Remember that one of the most common referral diagnoses is Cushing's syndrome, but this is one of the least common final diagnoses. When clinical suspicion is high, a screen for Cushing's syndrome is indicated.

The Screen for Cushing's Syndrome

Cushing's syndrome is the persistent oversecretion of cortisol. It can develop in 5 different ways: pituitary adrenocorticotropic hormone (ACTH) overproduction (Cushing's disease), ectopic ACTH overproduction by tumors, autonomous cortisol secretion by the adrenal or, very rarely, ovarian tumors, or a fifth extremely rare possibility, the secretion of corticotropin-releasing hormone (CRH) by a tumor. A clinician must first make the diagnosis of Cushing's syndrome (excessive cortisol secretion) before determining the etiology.

The most useful measurements in the basal state to detect Cushing's syndrome are the 24-hour urinary free cortisol excretion (10–90 µg) and the late evening plasma cortisol level (less than 15 µg/dL). The urinary excretion of 17-ketosteroids and 17-hydroxysteroids and measurement of morning and afternoon plasma cortisol levels are less reliable because of a significant overlap between normal and abnormal patients.

We prefer to start with the single-dose overnight dexamethasone test. Dexamethasone (1 mg) is given orally at 11 PM, and a plasma cortisol is drawn at 8:00 the next morning. A value less than 5 µg/dL rules out Cushing's syndrome. Cushing's syndrome is unlikely with intermediate values between 5 and 10 µg/dL, whereas a value higher than 10 µg/dL is diagnostic of adrenal hyperfunction. The number of patients with Cushing's syndrome who show a normal suppression in the single-dose overnight test is negligible (less than 1%). Obese patients, however, have a false-positive rate up to 13%.

If the single-dose overnight test is abnormal, establish the diagnosis by measuring the 24-hour urinary free cortisol. The low-dose, 2-day suppression test provides final confirmation. Dexamethasone (0.5 mg every 6 hours) is administered for 2 consecutive days after 2 days of baseline 24-hour urinary 17-hydroxysteroid and free cortisol measurements. Patients with Cushing's will not lower their urinary 17-hydroxysteroids below 2.5 mg/day and the free cortisol below 10 µg on the second day of dexamethasone suppression. Combining the low-dose test with the 24-hour urinary free cortisol should definitely provide the diagnosis of Cushing's syndrome. A 24-hour urinary free cortisol greater than 250 µg is virtually diagnostic of Cushing's syndrome, and a urinary free cortisol greater than 200 µg/day provides 90% diagnostic accuracy.
Pseudo-Cushing's states exist in patients with mild hypercortisolism due to conditions such as alcoholism, response to stress, anorexia and bulimia nervosa, severe obesity, and depression. Although usually not necessary, combining the low-dose dexamethasone suppression test with CRH stimulation is an accurate method to distinguish the real syndrome from the hypercortisolism associated with these other conditions. After two days of low-dose dexamethasone suppression, a single plasma cortisol level is obtained 15 minutes after administering CRH (1 μg/kg) intravenously. A 15-minute cortisol level greater than 1.4 μg/dL requires further evaluation.

The etiology of Cushing's syndrome can be established by combining a high dose dexamethasone suppression test with measurement of the basal state blood ACTH level. Dexamethasone (2 mg every 6 hours) is administered for 2 days, and the urinary 17-hydroxysteroid and cortisol levels on the second day are compared with basal levels. If basal ACTH is less than 5 pg/mL, and the urinary steroids do not decrease by at least 40%, an adrenal tumor is likely. When ACTH is measurable in the blood (more than 20 pg/mL), an ectopic ACTH-producing tumor is unlikely if the urinary steroids decrease by at least 40%. Cushing's disease is present when the blood ACTH level is in the normal range, a chest x-ray is normal, and imaging detects an abnormal sella turcica. A level of plasma ACTH greater than 50 pg/mL suggests ectopic ACTH release; a level less than 5 pg/mL suggests an autonomous cortisol-secreting tumor.

Imaging is very accurate and reliable in detecting adrenal tumors. In addition, imaging reliably predicts which patients have ectopic ACTH-producing tumors by detecting bilateral adrenal enlargement in such patients. CT scanning of the adrenal glands provides better resolution and is preferred over MRI and ultrasonography.

The evaluation of a patient with Cushing's syndrome can yield inconclusive results, and a failure to recognize an occult, ectopic ACTH-secreting neoplasm can lead to unnecessary pituitary or adrenal surgery. Bilateral venous sampling from the inferior petrosal sinus (sampling the blood draining from the pituitary gland) for the measurement of ACTH before and after CRH stimulation is an effective means to achieve accurate diagnosis of a pituitary origin for the ACTH. Approximately 15% of patients with ACTH-dependent Cushing's syndrome will have an occult, ectopic source for the ACTH. Most of these ACTH-secreting lesions are in the thorax (usually small-cell lung carcinoma), some in the abdomen. Petrosal sinus sampling is recommended in all patients with ACTH-dependent Cushing's syndrome who do not have an obvious adrenal tumor on imaging.

A very rare cause of Cushing's syndrome is the autonomous production of cortisol by an ovarian tumor. Chest and abdominal imaging is recommended for all atypical presentations.

Assessment of Insulin Secretion

Hyperandrogenism and hyperinsulinemia are commonly associated, as discussed in detail with complete references in Chapter 12. In many patients, a disorder in insulin action precedes the increase in androgens. Hyperinsulinemia can directly augment thecal cell androgen production in the ovary, and in addition, hyperinsulinemia contributes to the hyperandrogenism by inhibiting hepatic synthesis of sex hormone-binding globulin and insulin-like growth factor binding protein-1, actions that increase free testosterone levels and augment IGF-1 stimulation of thecal androgen synthesis, respectively.

Weight loss reduces both hyperinsulinemia and hyperandrogenemia, and is often followed by a return to ovulatory function. The use of oral agents such as troglitazone and metformin for the treatment of these patients offers great promise. Overweight, hyperandrogenic and hyperinsulinemic, anovulatory women must be counseled regarding their increased risk of future diabetes mellitus and cardiovascular disease.

The mechanism of hyperandrogenism can be attributed to hyperinsulinemia in an occasional case in which the circumstances are hard to understand; for example, the onset of hirsutism in an elderly woman who is found to have hyperthecosis in the ovaries. The appearance of hirsutism in such a case is not due to an ovarian response to hypergonadotropism but, rather, to the development of hyperinsulinemia.

For these reasons, we make the following recommendations:

All anovulatory women who are hyperandrogenic should be assessed for insulin resistance and glucose tolerance with measurements of:

1. The fasting glucose:insulin ratio, followed by
2. The 2-hour glucose level after a 75 g glucose load:
   - normal: less than 140 mg/dL
   - impaired: 140–199 mg/dL
   - noninsulin-dependent diabetes mellitus: 200 mg/dL and higher

In women who continue to manifest this disorder, periodic surveillance is necessary. The frequency is uncertain, but annual assessment is appropriate in women who continue to be obese.

Unfortunately, testing for hyperinsulinemia is not straightforward. Clinical guidelines for the interpretation of insulin levels and response to glucose loading have not been established. Because there is considerable overlap between normal women and patients with anovulation and polycystic ovaries, it is reasonable to assume that all overweight, anovulatory women with polycystic ovaries are hyperinsulinemic. Nevertheless, we recommend the measurement of the ratio of fasting glucose to fasting insulin in order to provide evidence that lends credence and importance to counseling efforts. A ratio of less than 4.5 is consistent with insulin resistance.

The DHAS Level

DHAS circulates in higher concentration than any other steroid and is derived almost exclusively from the adrenal gland. It is, therefore, a direct measure of adrenal androgen activity, correlating clinically with the urinary 17-ketosteroids. The upper limit of normal in most laboratories is 350 μg/dL, but, because of laboratory variation, attention must be paid to the local range of normal.

A random sample of DHAS is sufficient for the evaluation of hirsutism, needing no corrections for body weight, creatinine excretion, or episodic variation. Variations
Aging is associated with a decrease in the blood concentration of DHAS. The decrease accelerates after menopause, and DHAS is almost undetectable after age 70. This decline is 4 times greater than the age-related decline in cortisol, which is further support for the contention that cortisol and DHAS secretion are separately controlled.

Both 17-ketosteroids and circulating levels of DHAS are elevated in association with hyperprolactinemia. The levels return to normal with prolactin suppression by dopamine agonist treatment. In addition, increased free testosterone levels associated with decreased SHBG are often found in hyperprolactinemic women. This underscores the need to search for galactorrhea and to obtain a prolactin measurement in all anovulatory women. The androgen changes are probably secondary to the persistent anovulatory state induced by the elevated prolactin, although direct prolactin effects on the adrenal, ovary, or SHBG are possible.

When the DHAS level is normal, adrenal disease is most unlikely, and the diagnosis of excess androgen production by the ovaries is obvious. There are only rare cases of adrenal tumors with normal DHAS levels, and further evaluation of such cases would be indicated by the presence of markedly elevated blood levels of testosterone. These rare tumors are responsive to luteinizing hormone (LH), suggesting that they are derived from embryonic rest cells. Nonclassical (late-onset) adrenal hyperplasia commonly is not associated with an increased level of DHAS; the diagnosis of this condition relies on the measurement of 17-OHP for screening.

The clinical problem with DHAS measurements in the evaluation of hirsutism is the frequent finding of a moderately elevated DHAS level in anovulatory patients with polycystic ovaries. This is similar to the moderate elevations of 17-ketosteroids encountered in these patients. If the 17-OHP level is normal, we believe that it is worthwhile to subject these patients to a search for an adrenal enzyme defect. Clinical experience has established that moderate elevations of DHAS are associated with anovulation; suppression of ovarian function restores the DHAS level to normal.

A DHAS level of 700 µg/dL or greater has been accepted as a marker for abnormal adrenal function. But how often is a DHAS of this level encountered? DHAS levels of 700 µg/dL or more are confronted so rarely that we must now question the clinical usefulness of measuring DHAS. We cannot identify a single case in our experience in which the DHAS level changed patient management. Even if a DHAS of 700 µg/dL or greater is encountered, we believe that the adrenal secretion of extremely high levels of DHAS will be associated with high levels of testosterone, either by direct secretion or by peripheral conversion of the DHAS. Therefore, in the absence of Cushing’s syndrome, we believe that measurement of testosterone will suffice to screen for adrenal abnormalities. An imaging assessment of the adrenal glands whenever a markedly elevated testosterone is encountered will be more cost-effective than measuring DHAS in all hirsute women.

Nonclassical Adrenal Hyperplasia

Congenital adrenal hyperplasia is due to an enzyme defect leading to excessive androgen production. This severe condition, with its prenatal onset, is inherited in an autosomal recessive fashion (discussed in Chapter 9). A milder form of the disease, appearing later in life, has been designated by a variety of adjectives, including late-onset, partial, nonclassical, attenuated, and acquired adrenal hyperplasia. An asymptomatic form, cryptic adrenal hyperplasia, is revealed only with biochemical testing.

Although each of the enzymatic steps from cholesterol to cortisol can be affected in specific clinical disease, the most common enzymes to be deficient are 21-hydroxylase and 3β-hydroxysteroid dehydrogenases. Unlike 21-hydroxylase deficiency, no genetic markers are currently available; diagnosis requires ACTH stimulation and demonstration of an altered 17α-hydroxyprogrenolone to 17-OHP ratio. Some argue that late-onset 3β-hydroxysteroid dehydrogenase deficiency is encountered more frequently than 21-hydroxylase deficiency. However, although an exaggerated 17α-hydroxyprogrenolone response to ACTH stimulation is common in women with hyperandrogenism, the response is consistent with adrenal hyperactivity and not an enzyme deficiency. Furthermore, molecular studies fail to find mutations in the genes for the two 3β-hydroxysteroid dehydrogenase enzymes in patients who appear to have mild to moderate deficiencies in 3β-hydroxysteroid dehydrogenase. We believe that this deficiency is so subtle that accurate diagnosis is not essential. Our usual therapeutic approach to hirsutism will be effective. The 11β-hydroxylase deficiency is quite rare, and it is usually diagnosed at a younger age (Chapter 9). It is not worth measuring the 11-deoxycorticisol response to ACTH stimulant in adult hirsute women in an effort to detect this rare deficiency.

The 21-Hydroxylase Defect

Women with late-onset adrenal hyperplasia due to a 21-hydroxylase deficiency respond to ACTH stimulation in a moderate fashion, between the classical homozygote response and the mild heterozygote reaction (see the 17-OHP nomogram). A 21-hydroxylase deficiency is now recognized to be the most common autosomal recessive disorder, surpassing cystic fibrosis and sickle cell anemia. The clinical presentation is extremely variable, and the symptoms may appear and disappear over time. Therefore, the diagnosis requires laboratory evaluation as discussed under “The 17-OHP Level.” The genetic diagnosis for known mutations in the CYP21 gene is discussed in Chapter 9.

There are at least 3 reasons that make it worthwhile to seek the correct diagnosis:

1. Therapy should be accurately applied because it must be long-term.
2. Pregnant couples with this condition require genetic counseling for the prenatal diagnosis and possible treatment of the congenital form of the disease, as well as the assessment of asymptomatic offspring. However, without knowing the father's carrier status, an accurate estimate of risk is impossible. Although the risk of having a child with congenital adrenal hyperplasia is very low, the couple should consider prenatal testing for heterozygosity. If the father tests positively, prenatal diagnosis and treatment would be reasonable.
3. Theoretically, these patients might be subject to cortisol deficiency during severe stress; however, to our knowledge this has not been a clinical problem.

Other Enzyme Defects

The 3β-hydroxysteroid (type II) dehydrogenase deficiency exists in both the ovaries and adrenals. This defect precludes significant androgen production; however, this enzyme activity appears to remain intact in peripheral tissues. Therefore, hirsutism seen with this deficiency is probably due to target tissue conversion of the increased secretion of precursors. Unlike 21-hydroxylase deficiency, no genetic markers are currently available; diagnosis requires ACTH stimulation and demonstration of an altered 17α-hydroxyprogrenolone to 17-OHP ratio. Some argue that late-onset 3β-hydroxysteroid dehydrogenase deficiency is encountered more frequently than 21-hydroxylase deficiency. However, although an exaggerated 17α-hydroxyprogrenolone response to ACTH stimulation is common in women with hyperandrogenism, the response is consistent with adrenal hyperactivity and not an enzyme deficiency. Furthermore, molecular studies fail to find mutations in the genes for the two 3β-hydroxysteroid dehydrogenase enzymes in patients who appear to have mild to moderate deficiencies in 3β-hydroxysteroid dehydrogenase. We believe that this deficiency is so subtle that accurate diagnosis is not essential. Our usual therapeutic approach to hirsutism will be effective. The 11β-hydroxylase deficiency is quite rare, and it is usually diagnosed at a younger age (Chapter 9). It is not worth measuring the 11-deoxycorticisol response to ACTH stimulant in adult hirsute women in an effort to detect this rare deficiency.

The 17-OHP Level

From 1 to 5% of women who complain of hirsutism display a biochemical response that is consistent with the less severe form of adrenal hyperplasia of the 21-hydroxylase variety. This relative frequency of late-onset adrenal hyperplasia dictates routine 17-OHP screening of women who complain of hirsutism. On the other hand, the routine use of the ACTH stimulation test is not warranted. Heterozygosity for CYP21 mutations does not increase the risk of clinically significant hirsutism.

Besides using the 17-OHP screen to make a cost-effective decision regarding ACTH stimulation, one can be swayed by pertinent clinical findings. A strong family history of androgen excess suggests the presence of an inherited disorder. Hirsutism due to an adrenal enzyme defect usually is more severe and begins at a young age, typically at puberty. Short stature and very high blood levels of androgens also signify a more severe problem. Finally, it is worth considering the following: With normal baseline steroid levels, even if a woman has a subtle enzyme defect, the management of the problem does not require its discovery.
17-OHP must be measured first thing in the morning to avoid later elevations due to the diurnal pattern of ACTH secretion. The baseline 17-OHP level should be less than 200 ng/dL. Levels greater than 200 ng/dL, but less than 800 ng/dL, require ACTH testing. Levels over 800 ng/dL are virtually diagnostic of the 21-hydroxylase deficiency. The DHAS level is usually normal. The hallmarks of late-onset adrenal hyperplasia are elevated levels of 17-OHP and a dramatic increase after ACTH stimulation. However, the elevated baseline levels of 17-OHP are often not impressive (e.g., overlapping with those found in women with polycystic ovaries due to anovulation), and a simple ACTH stimulation test must be utilized.

17-OHP must be measured first thing in the morning to avoid later elevations due to the diurnal pattern of ACTH secretion. The baseline 17-OHP level should be less than 200 ng/dL. Levels greater than 200 ng/dL, but less than 800 ng/dL, require ACTH testing. Levels over 800 ng/dL are virtually diagnostic of the 21-hydroxylase deficiency. The DHAS level is usually normal. The hallmarks of late-onset adrenal hyperplasia are elevated levels of 17-OHP and a dramatic increase after ACTH stimulation. However, the elevated baseline levels of 17-OHP are often not impressive (e.g., overlapping with those found in women with polycystic ovaries due to anovulation), and a simple ACTH stimulation test must be utilized.

The ACTH Stimulation Test

Synthetic ACTH (Cortrosyn) is administered intravenously in a dose of 250 µg. Blood samples for the measurement of 17-OHP are obtained at time 0 and again at 1 hour. The testing must be performed in the morning (8 AM), but it can be scheduled at any time during the menstrual cycle. The 1 hour value is plotted on the nomogram which predicts the genotype of homozygote and heterozygote forms of the 21-hydroxylation deficiency. Dexamethasone pretreatment the night before is not necessary. Heterozygote carriers for 21-hydroxylation deficiency have ACTH-stimulated levels of 17-OHP up to 1000 ng/dL; patients with late-onset deficiency have stimulated levels above 1200 ng/dL.

For the diagnosis of the 3b-hydroxysteroid dehydrogenase deficiency, the same ACTH stimulation test is utilized, measuring 17-OHP and 17-hydroxypregnenolone. An abnormal 17-hydroxypregnenolone:17-OHP ratio is usually greater than 6.0. This deficiency is also usually marked by a significant elevation of DHAS in the face of normal or mildly elevated testosterone levels. In the 11b-hydroxylation deficiency, the level of 11-deoxycorticisol will be increased; it is normal with the 21-hydroxylation defect.

The Adrenal Gland and Anovulation

Adrenal involvement in the syndrome of anovulation and hirsutism has long been recognized. Adrenal suppression, for example, will induce regular menses and ovulation in some patients, and empiric treatment with glucocorticoids has been advocated in the past.

Late-onset adrenal hyperplasia does not explain every anovulatory woman encountered with a moderate elevation in DHAS. The important clinical question is the following: is excessive androgen secretion by the adrenal gland a primary disorder in these women, or is it a secondary reaction to the hormonal milieu associated with anovulation?

One possibility is that the adrenal hyperactivity (as indicated by elevated DHAS levels) is due to an estrogen-induced 3b-hydroxysteroid dehydrogenase insufficiency. Considerable effort has been devoted to demonstrating an estrogen influence on adrenal androgen secretion. Unfortunately there is no clear-cut conclusion, with both positive and negative results reported.

This picture may be similar to that of the fetal adrenal gland. Studies have demonstrated that the low level of 3b-hydroxysteroid dehydrogenase activity and high secretory levels of DHAS by the fetal adrenal cortex are due to estrogen. Inconsistent with this explanation is the fact that ACTH levels in adult anovulatory women are not elevated; however, the period of increased ACTH response would exist only until normal cortisol levels are re-achieved (a new set point). Indeed, studies of anovulatory women with elevated levels of DHAS indicate that the increase in adrenal activity is because of a mechanism within the adrenal gland, not because of an increase in pituitary response to CRH and not because of increased adrenal response to ACTH stimulation.

The activity of 3b-hydroxysteroid dehydrogenase is inhibited by both androgens and estrogens in concentrations to be expected within the adrenal gland but difficult to achieve with exogenous administration; changes in adrenal secretion, therefore, can reflect varying action of steroids, especially estrogen, in different layers of the adrenal cortex. The failure to affect adrenal steroidogenesis with the exogenous administration of androgens is consistent with this hypothesis. Thus, it remains attractive to explain adrenal hyperactivity seen in anovulatory women as a secondary reaction induced and maintained by the constant estrogen state associated with persistent anovulation. Indeed, in patients with polycystic ovaries and increased adrenal androgen activity, a correlation can be demonstrated between adrenal sensitivity to ACTH and estrogen levels. This relationship, however, may be the result of mechanisms other than inhibition of 3b-hydroxysteroid dehydrogenase (see discussion of the fetal adrenal gland in Chapter 8).

Suppression of ovarian function by treatment with a GnRH agonist has been utilized in hopes of bringing clarity to this puzzle by assessing adrenal function after elimination of ovarian steroid production. Short-term suppression (3–6 months) has been reported to have no impact on adrenal androgen production. However, these studies did not include women with high levels of DHAS. When anovulatory women with levels of DHAS greater than normal are treated with a gonadotropin-releasing hormone (GnRH) agonist for at least 3 months, in some women, but not all, the elevated DHAS levels are suppressed. Thus, the exact
nature of the adrenal-ovarian relationship in these women may vary, consistent with the clinical heterogeneity characteristic of these patients. It has been suggested that increased adrenal androgen production in some women is secondary to P450c17 17,20 lyase hyperactivity, and in others it is an acquired response due to the anovulatory hormonal steady state. In those women with P450c17 hyperactivity, the underlying disorder may be hyperinsulinemia. Regardless of differences in weight, diet, race, and environmental factors, excess adrenal androgen activity is present in one-half to two-thirds of anovulatory women and hyperinsulinemia in approximately 70%. It makes sense that similar growth factor modulation takes place in adrenal steroid-producing cells as in the ovary. Women with polycystic ovaries (both obese and nonobese) and hyperinsulinemia have a greater steroidogenic response to ACTH than anovulatory patients with normal insulin levels. Insulin and IGF-I receptors are present on adrenal cells. The infusion of insulin into women causes a decrease in DHAS, and hyperinsulinemia inhibits adrenal 17,20 lyase (P450c17) activity, suggesting that insulin decreases the production of this adrenal androgen. It is further suggested that the age-related decline in DHAS may be related to increasing insulin resistance. How do these changes relate to DHAS levels in anovulatory women, levels that range from normal to moderately elevated? A simple inverse relationship of adrenal androgen levels in anovulatory women does not exist. Furthermore, in a study of 92 consecutive women with hirsutism, the steroidogenic response to ACTH stimulation was not consistent with an inherent disorder of P450c17.

An ovulatory response following the treatment of anovulatory women with dexamethasone can be explained in part by a contribution made to ovarian androgen production by circulating DHAS. A significant percentage of testosterone production by the ovarian follicle can be attributed to circulating DHAS serving as a substrate or prehormone. Thus, dexamethasone suppression cannot separate adrenal and ovarian testosterone secretion in that the two glands are involved in a complex interaction, with DHAS providing at least one mechanism for this interaction.

The inconclusiveness of this situation and the rarity of a true adrenal enzyme deficiency in adult women make a cost-effective argument against routine endocrine testing. Accordingly, we have adopted a 17-OHP level of 200 ng/dL, below which we do not pursue the possibility of a primary adrenal enzyme problem, and we now find little use for DHAS measurements. Mild adrenal enzyme defects can be treated by our usual methods and do not require glucocorticoid administration.

The Testosterone Level

Plasma testosterone levels (normal 20–80 ng/dL) are elevated in the majority of women (70%) with anovulation and hirsutism. Individual variation is great, however, largely because of the changes in the testosterone-binding capacity of the sex hormone-binding globulin in the blood. Because the binding globulin levels are depressed by androgen and insulin, the total testosterone concentration can be in the normal range in a woman who is hirsute even though the percent unbound and active testosterone is elevated. Indeed, the unbound or free testosterone is approximately twice normal (an increase from 1% to 2%) in women with anovulation and polycystic ovaries. Therefore, a normal total testosterone level in hirsute women is still consistent with elevated androgen production rates.

It is not necessary to measure the free testosterone (a technically difficult and expensive assay) because a routine total testosterone assay adequately serves the purpose of screening for testosterone-secreting tumors. Such tumors are associated with testosterone levels that are usually in the male range, and, therefore, the fine discrimination of the free testosterone level is unnecessary. If the testosterone level exceeds 200 ng/dL, an androgen-producing tumor must be suspected.

The arbitrary cutoff point has been challenged because variations in secretion can yield misleading values, and not all androgen-producing tumors will be this active, and some women with polycystic ovaries (especially hyperthecosis) will have testosterone levels greater than 200 ng/dL. Nevertheless, the combination of the patient’s historical chronology of the development of hirsutism, the pelvic examination, and the 200 ng/dL testosterone level will provide accurate diagnosis in virtually all cases. A patient with an acute, rapid course of virilizing symptoms requires a full evaluation for the presence of an androgen-producing tumor even if the testosterone concentration is less than the cutoff level.

Be aware that testosterone levels are significantly elevated during normal pregnancy. Levels are greater than 100 ng/dL in the first trimester and reach 500 to 800 ng/dL by term. This is mainly due to the estrogen-induced increase in sex hormone-binding globulin; however, even the free testosterone level is elevated in the third trimester. Mother and fetus are protected from these high androgen levels by many mechanisms, including binding to sex hormone-binding globulin and placental aromatization of androgens to estrogens.

Androgen-Producing Tumors

There are two findings that should stimulate the clinician to suspect the presence of an androgen-producing tumor. One is a history of rapidly progressive masculinization. Hirsutism associated with anovulation is generally slow to develop, usually covering a time period of at least several years. Tumors are associated with a short time course, measured in months. The second finding that should arouse suspicion is a testosterone level greater than 200 ng/dL.

In our view, androgen-producing tumors are one of medicine’s vastly overrated problems. First, they are incredibly rare, yet they attract an inordinate amount of attention at our meetings and a disproportionate number of printed pages in texts and journals. Second, there is an endocrine mystique surrounding the functioning tumor. Actually it is a straightforward problem.

Functioning ovarian tumors are almost all palpable, and like any ovarian mass, rapid laparotomy and surgical removal are in order. It is well recognized, however, that very small ovarian tumors (usually in the hilus of the ovary) can secrete testosterone. And occasionally, virilization is encountered with a nonfunctional tumor due to tumor stimulation of androgen secretion in the surrounding stromal tissue.

The only diagnostic dilemma is when to explore the patient in whom a mass is not palpable. Suppression and stimulation tests are known to falsely lead to oophorectomy in the presence of a virilizing adrenal adenoma. In addition, suppression and stimulation methods do not specifically isolate ovarian or adrenal function. Ovarian androgen-producing tumors are responsive to LH and, therefore, will respond to ovarian suppression and stimulation. Selective angiography with venous sampling and measurement of adrenal and ovarian steroids is not without problems. It is technically difficult to achieve bilateral catheterization of the ovaries, steroid secretion is episodic (especially by the adrenal glands), and the technique is not without risk. Selective retrograde catheterization of ovarian and adrenal veins by an expert should be reserved for those few patients who have been imaged with negative findings and yet the clinical
Incidentally discovered adrenal masses require evaluation. A clinician should always consider the possibility of hyperinsulinemia and emphasize preventive health interventions (as discussed in Chapter 30).

The single-dose overnight dexamethasone test is used to screen for Cushing’s syndrome. Abnormal results are confirmed by measuring the 24-hour measurements among the various products.

In postmenopausal women with hyperandrogenism, it is usually appropriate to be more aggressive surgically; however, keep in mind that hyperinsulinemia in the postmenopausal years can stimulate hyperandrogenism, which would simulate the presentation of a tumor. GnRH agonist treatment can avoid surgery in these patients because insulin-induced steroidogenic activity in the ovary is still LH-dependent.

When an androgen-producing tumor is suspected and an adrenal mass is not palpable, imaging of the adrenal glands and ovaries should be obtained. Imaging of the adrenal is a sensitive diagnostic technique for small tumors that produce Cushing’s syndrome as well as for virilizing adrenal adenomas. For adrenal imaging, CT scanning provides better resolution and is preferred over MRI and ultrasonography. For ovarian imaging, transvaginal ultrasonography is the method of choice.

### The Incidental Adrenal Mass

Adrenal masses will be discovered incidentally in approximately 10% of postmortem examinations, and therefore, it is to be expected that an incidental adrenal mass will be discovered occasionally with abdominal imaging. Bilateral lesions are more serious. Common causes of bilateral lesions are metastatic cancer (most commonly from breast, kidney, or lung), infection (tuberculous and fungal), and adrenal hyperplasia; hence, surgery is rarely indicated. A primary malignancy of the adrenal is usually associated with excess secretion of glucocorticoids and androgens. The size of a lesion is significant. The probability of malignancy roughly parallels the diameter of the lesion. Bilateral lesions less than 3 cm usually are due to metastatic disease. Thus, the current recommendation is to excise unilateral masses if they are of significant size, usually greater than 4 cm in diameter. Fine needle aspiration is also recommended for all unilateral adrenal lesions, after excluding a pheochromocytoma, to rule out metastatic lesions.

When following a mass, imaging should be performed at 3, 9, and 18 months. Any mass that is stable after 18 months can be left in place.

Incidental adrenal masses require evaluation for biochemical function. The presence of hypertension raises the suspicion of Cushing’s syndrome, hyperaldosteronism, or pheochromocytoma. The evaluation should include a screening test for pheochromocytoma (24-hour urinary catecholamines), electrolyte, aldosterone, and renin activity assessment for aldosteronism, a 24-hour urinary free cortisol, and a testosterone level. If these tests are normal, subclinical hormone secretion can be present. A relatively high incidence of cortisol-secreting tumors in the presence of normal 24-hour urinary free cortisol excretion indicates the need to perform an overnight dexamethasone suppression test in patients with asymptomatic incidental adrenal masses. There appears to be a high incidence of adrenal masses in patients with adrenal hyperplasia, presumably because of chronic excess ACTH stimulation. These need not be removed surgically, but a laboratory evaluation for adrenal hyperplasia is, therefore, indicated in patients with incidental adrenal masses to avoid unnecessary surgery. Finally, the clonidine suppression test is useful to diagnose a subclinical pheochromocytoma. Adrenal scintigraphy uses radiotracers accumulated by functioning tissue and is recommended when biochemical testing is normal. When uptake in the mass is less than in the contralateral adrenal (a “discordant” pattern), fine-needle aspiration is indicated to rule out a malignant lesion.

#### Screening Tests for Incidental Adrenal Masses

- 24-hour urinary catecholamines and free cortisol
- Testosterone
- Renin activity, aldosterone, and electrolytes

#### Provocative Tests for Subclinically Active Incidental Adrenal Masses

- Dexamethasone overnight suppression test
- 17-Hydroxyprogesterone response to ACTH
- Clonidine suppression test (clonidine, 0.3 mg p.o in a supine position, followed by plasma norepinephrine levels at 0, 2, and 3 hours; a norepinephrine level above 500 pg/mL or 50% greater than the 0-hour level is a positive result)

### Summary of Key Recommendations for the Evaluation of Hirsutism

1. The laboratory evaluation of hirsutism consists of the measurement of the circulating levels of testosterone and 17-OHP. When alopecia is present, a TSH screen for thyroid function is also indicated.
2. The single-dose overnight dexamethasone test is used to screen for Cushing’s syndrome. Abnormal results are confirmed by measuring the 24-hour urinary free cortisol.
3. A clinician should always consider the possibility of hyperinsulinemia and emphasize preventive health interventions (as discussed in Chapter 12).
4. Any patient with rapidly progressive virilization must be evaluated for an androgen-secreting tumor regardless of the results of screening laboratory tests.
5. Incidentally discovered adrenal masses require evaluation.

### Treatment of Hirsutism

Almost all patients presenting with hirsutism represent excess androgen production in association with the steady state of persistent anovulation. Treatment is directed toward interruption of the steady state. In those patients who wish to become pregnant, ovulation can be induced as discussed in Chapter 30. In patients who do not want to become pregnant, the steady state can be interrupted by suppression of ovarian steroidogenesis by utilizing the potent inhibitory action on LH of gestational agents.

Androgen production in hirsute women is usually an LH-dependent process. Suppression of ovarian steroidogenesis depends upon adequate LH suppression. In addition to the inhibitory action of the gestational component, oral 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 oral contraceptives also inhibit 5a-reductase activity in skin, further contributing to the clinical impact of oral contraceptives on hirsutism.

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. The beneficial clinical effect is the same with low-dose preparations containing levonorgestrel, previously recognized to cause acne at high dosage. 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. Theoretically, these products would be more effective in the treatment of acne and hirsutism; however, this is not 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).

Even in women being treated with antiandrogens, oral contraceptives are important and useful to provide both cycle control and contraception.
In the patient in whom oral contraceptives are contraindicated or unwanted, good results can be achieved with the use of medroxyprogesterone acetate, either 150 mg intramuscularly every 3 months or 10–20 mg orally per day. The mechanism of action of medroxyprogesterone acetate is slightly different from that of the combination oral contraceptive. Suppression of gonadotropins is less intense; hence, ovarian follicular activity continues. LH suppression is significant, however, and testosterone production is decreased, although to a lesser degree than with combined oral contraceptives. In addition, testosterone clearance from the circulation is increased. This latter effect is due to an induction of liver enzyme activity. Medroxyprogesterone acetate decreases SHBG so that less testosterone is bound; however, suppression of total testosterone production is so great that the actual amount of free testosterone decreases. The overall effect yields a clinical result comparable with that achieved with the combination oral contraceptive.

A noteworthy feature of hirsutism is the slow response to treatment. Because of the hair growth cycle, change takes time. The patient should be cautioned that treatment with hormonal suppression will be necessary for at least 6 months before an observable diminution in hair growth occurs. Combined treatment with electrolysis is not recommended, therefore, until hormonal suppression has been used at least 6 months.

New hair follicles will no longer be stimulated to grow, but hair growth that has been previously established will not disappear with hormone treatment alone. This can be affected temporarily by shaving, tweezing, waxing, or the use of depilatories. Contrary to a common belief, these methods do not stimulate growth or increase the rate of growth of hair. None of these tactics alters the inherent growth of the hair; therefore, they must be reapplied at frequent intervals. Permanent removal of hair can be accomplished only by electrocoagulation of dermal papillae. Patients should be counseled to make sure their electrologists use disposable needles.

Some patients return after a period of treatment expressing disappointment because hair is still present. The effect of the treatment (prevention of new hair growth) may not be apparent unless the previously established hair is removed. The combination of ovarian suppression preventing new hair growth and electrolysis removing the old hair yields the most complete and effective treatment of hirsutism.

How long should treatment be continued? After 1–2 years, it is worthwhile to stop the medication and observe the patient for a return of ovulatory cycles. Even in those patients who continue to be anovulatory, testosterone suppression continues for 6 months to 2 years after discontinuing treatment. Of course, if anovulation is still present, one can expect the eventual return of hirsutism.

The really resistant patient deserves further consideration. Combination therapy with one of the methods discussed below is worthwhile, preferably an oral contraceptive combined with either spironolactone or finasteride.

In most patients, DHAS levels are suppressed by progestational treatment. The mechanism is not definitely known, but there are several possible explanations. If the original stimulus for the increased DHAS secretion is the steady estrogen state of anovulation, then the change in the endocrine milieu of the adrenal gland brought about by the suppression of ovarian steroidogenesis will restore a normal adrenal secretory pattern. Oral contraceptives may also produce subtle but significant alterations in ACTH secretion or response in the adrenal gland.

The effectiveness of adrenal suppression in inducing ovulatory cycles in some anovulatory patients can be attributed to a lowering of circulating androgen levels due to a decrease in the adrenal contribution as well as a reduction in the amount of DHAS available for conversion to testosterone within the ovarian follicle. The intraovarian androgen level is decreased, therefore lowering the inhibitory action of androgens on follicular growth and development. In terms of ovulation, the frequency of successful response with this type of treatment does not match that of the first drug of choice, clomiphene. In terms of treatment of hirsutism, progestin suppression of ovarian steroidogenesis is more effective and should remain the first therapeutic approach. Adrenal suppression should be reserved for patients with a clearly established diagnosis of an adrenal enzyme deficiency.

In an older woman who has no further desire for fertility, and in the woman for whom continued use of steroid medication is disturbing because of increasing risks with increasing age, serious consideration should be given to a surgical solution. A persistent problem of hirsutism, especially if it is progressive in severity, is a reasonable indication for hysterectomy and bilateral salpingo-oophorectomy. Patients with hyperthecosis respond poorly to suppression and are usually older. Surgical treatment for these patients is often very appropriate. Of course, an estrogen regimen is recommended for these patients postoperatively.

Keep in mind the strong association between hyperandrogenism and hyperinsulinemia as discussed in Chapter 12. Treating the problem of hirsutism by suppression of androgen production and action will not restore glucose metabolism to normal; these patients will continue to demonstrate insulin resistance. In overweight, hyperandrogenic patients, a weight control program is essential to reduce the risks of diabetes mellitus and cardiovascular disease, and consideration should be given to the preventive health benefits expected from long-term treatment with metformin or troglitazone. Nevertheless, hirsutism in hyperandrogenic women with hyperinsulinemia will respond favorably to treatment with oral contraceptives.

**Additional Methods of Treatment**

**Spironolactone**

Spironolactone is an aldosterone-antagonist diuretic. In the treatment of hirsutism, spironolactone has multiple actions, inhibiting the ovarian and adrenal biosynthesis of androgens, competing for the androgen receptor in the hair follicle, and directly inhibiting 5a-reductase activity. The inhibition of steroidogenesis is achieved through an effect on the cytochrome P450 system, but the steroid suppressive effects are so variable that the receptor-blocking action is the most important mechanism. It is probably for this reason that cortisol, DHA, and DHAS levels are not significantly changed with spironolactone treatment, even though
The impact of spironolactone treatment on hirsutism is related to dosage, and a better effect is seen with a dose of 200 mg daily. After a period of time, one can usually lower the dose of spironolactone to a maintenance dose of 25–50 mg daily. As with progesterational agents, the response is relatively slow, and a maximal effect can be demonstrated only after 6 months of treatment. Side effects are minimal, including diuresis in the first few days of use, occasional complaints of fatigue, ad dysfunctional uterine bleeding. Remember that the anovulatory state requires progesterational management in order to avoid abnormal uterine bleeding (and endometrial hyperplasia). Because of the possibility of hyperkalemia, spironolactone should be used with caution in women who are elderly, diabetic, or using drugs that raise the potassium level.

We use spironolactone when patients find oral contraceptives unacceptable or the response is disappointing. Indeed, it makes sense to combine the peripheral tissue action of spironolactone with oral contraceptives to achieve a more dramatic result; however, the results with combined treatment regimens have not been impressively better than single agent therapy. Acne has been effectively treated with the local application of a cream containing 2–5% spironolactone. Systemic absorption does not occur, and there are no side effects.

One word of caution: with inhibition of androgen secretion, ovulation can occur, and effective contraception is important. Theoretically, spironolactone interference with testosterone action could result in the feminization of a male fetus; however, a disrupting effect on the external genitalia has not occurred, despite fetal exposure to high doses. Combined treatment with an oral contraceptive may produce a better clinical effect and, at the same time, prevent menstrual irregularity and provide contraception.

Cyproterone Acetate

Cyproterone is a potent progesterational agent that both inhibits gonadotropin secretion and blocks androgen action by binding to the androgen receptor. In many parts of the world, it has been used as an oral contraceptive agent called "Diane" (2 mg cyproterone acetate and 50 µg ethinyl estradiol). "Dianette" or "Diane 35" contains 2 mg cyproterone acetate and 35 µg ethinyl estradiol. In a method for the treatment of hirsutism called the reversed sequential regimen, cyproterone acetate is given in a dose of 50 or 100 mg daily on days 5–14, with 30 or 50 µg of ethinyl estradiol daily on days 5–25. In a comparison of Diane with the high dose (100 mg) cyproterone acetate treatment, the therapeutic effect was greater (but probably not of clinical significance) with the higher dose, and there was a similar incidence of side effects with both treatments. In a comparison of Dianette with higher doses of cyproterone acetate (20 and 100 mg), the clinical response with the 2 mg dose of cyproterone in Dianette was equal to that of the higher doses. The most common reactions include fatigue, edema, loss of libido, weight gain, and mastalgia. Significant improvement in facial hirsutism is seen by the 3rd month of treatment. In comparisons of spironolactone and cyproterone acetate, a monophasic low-dose oral contraceptive combined with spironolactone 100 mg daily was as effective as the reversed sequential regimen of 50 or 100 mg cyproterone and estrogen.

Dexamethasone

Dexamethasone suppression of endogenous ACTH secretion is used in women who have an adrenal enzyme deficiency. Dexamethasone is given nightly (to achieve maximal suppression of the central nervous system-adrenal axis that peaks during sleep) in a dose of 0.5 mg. An equivalent dose of prednisone is 5–7.5 mg. If this treatment suppresses the morning plasma cortisol level below 2.0 µg/dL, the dose should be reduced to avoid an inability to react to stress. Fortunately, adrenal androgen secretion is more sensitive to suppression by dexamethasone than is cortisol secretion. Patients with adrenal hyperplasia may require higher doses to normalize the steroid blood levels. With higher doses, alternate day therapy can still accomplish significant adrenal androgen suppression without affecting cortisol secretion. It should be emphasized that moderate elevations of DHA do not indicate patients who will benefit from dexamethasone treatment. Maximal effectiveness against hirsutism in patients with an adrenal enzyme deficiency may require treatment besides glucocorticoid supplementation. The addition of an oral contraceptive or antiandrogen should be considered.

Treatment With GnRH Agonists

Because ovarian androgen production is LH-dependent, suppression of the pituitary with chronic GnRH agonist treatment improves hirsutism. However, inconsistent results in the literature attest to the fact that sufficient dosage must be administered to achieve effective suppression and clinical response. Therefore, monitoring dosage and response is recommended. A greater effect of GnRH agonist is required to suppress ovarian androgen secretion compared with estradiol secretion. We recommend the use of depot administration of a GnRH agonist with monitoring treatment by measuring testosterone levels (the goal being less than 40 ng/dL). To avoid the problems associated with estrogen deficiency, estrogen-progestin add-back should be initiated after the maintenance dose of a long-acting GnRH agonist has been established. We recommend the daily administration of 0.625 mg conjugated estrogens or 1.0 mg estradiol combined with 2.5 mg medroxyprogesterone acetate or 0.35 mg norethindrone, or, even better, an oral contraceptive.

Although a combination of a GnRH agonist and an oral contraceptive produces a greater reduction in free testosterone levels, improvement of hirsutism is similar, comparing the combination to treatment with only an agonist; however, treatment with an agonist has been reported to be more effective than oral contraceptives alone. In one clinical trial, the improved response associated with agonist treatment was observed only in the first 3 months of treatment; by 6 months the combination of oral contraceptives and a GnRH agonist was no more effective than oral contraceptives. The addition of a long-acting agonist to Dianette (2 mg cyproterone acetate and 35 µg ethinyl estradiol) did not significantly improve clinical results. After one year, a comparison of GnRH agonist treatment with the high dose cyproterone acetate regimen indicated equal efficacy, although agonist treatment was followed by a more prolonged remission. These mixed results reflect variability in the severity of the condition and the degree in androgen suppression. The impact of a GnRH agonist-oral contraceptive combination is maximal when testosterone levels are suppressed below 40 ng/dL, and in patients who are overweight with severe hirsutism.

This method of treatment is relatively complicated and expensive and should be reserved for the severe case of ovarian hyperandrogenism, which is usually due to significant hyperinsulinism and elevated thyroid hormone (a condition that responds less well to the usual methods of treatment). An alternative that deserves consideration is treatment of the hyperinsulinemia with metformin or troglitazone. Once maximal response has been obtained with one of these more expensive methods, long-term suppression of hair growth can be maintained with an oral contraceptive or an antiandrogen.

Flutamide

Flutamide (Eulexin) is a nonsteroidal antiandrogen at the receptor level. Flutamide directly inhibits hair growth without many side effects (dry skin is the most common); however, hepatotoxicity is possible. Because of the uncommon but severe toxic effect on the liver, a low-dose approach is recommended. A dose of 250
mg daily can have a marked beneficial impact on hirsutism within 6 months. Despite the use of lower doses, however, it is prudent to monitor liver enzymes. In a comparison study, flutamide (250 mg bid) was not more effective than spironolactone (100 mg/day). Treatment with flutamide should be combined with a method of contraception; blockage of androgen receptors in a male fetus could interfere with normal male development. In our view, the potential for hepatotoxicity makes flutamide an unsatisfactory choice for treatment of hirsutism.

Finasteride

Finasteride inhibits 5a-reductase activity, thus blocking conversion of testosterone into dihydrotestosterone. The 5a-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 predominately expressed in reproductive tissues. Finasteride (Proscar), used to treat prostate cancer, effectively inhibits both isozymes, and, therefore, can be used to treat hirsutism. A dose of 5 mg per day decreases hirsutism without side effects. A lower dose is probably as effective. Finasteride has been available only as a 5 mg unscored tablet, but it can with care be cut into quarters. A smaller dose, 1 mg (Proppecia), has now been approved for the treatment of hair loss in men. In a randomized, clinical trial finasteride and spironolactone (100 mg daily) were equally effective. In another randomized clinical trial, spironolactone in a dose of 100 mg daily was more effective than finasteride. The main advantage of finasteride is the lack of side effects. Because the development of the urogenital sinus and urogenital tubercle into the male external genitalia, urethra, and prostate requires the action of dihydrotestosterone, patients being treated with finasteride should be cautioned regarding this possible risk during pregnancy, and an effective method of contraception must be used.

Other Agents

Cimetidine (300 mg qid) has been used to treat hirsutism, but it is the least potent of the androgen receptor blockers, and the clinical response is disappointing. The use of a skin cream containing progesterone is effective, but it must be applied frequently (because of rapid metabolic clearance) and its action is very concentrated at the point of application. Minoxidil in a topical application produces a moderate increase in hair growth in women with alopecia; however, indefinite, on-going treatment is necessary, and a return to the previous hair pattern cannot be achieved. Ketoconazole in a dose of 400 mg daily blocks androgen synthesis by inhibiting the cytochrome P450 system. Although the impact on hirsutism is significant, there is a high incidence of side effects as well as changes in liver enzymes. Ketoconazole, at best, should be a last resort, requiring frequent monitoring of liver function. In addition, chronic treatment with ketoconazole may suppress adrenal corticosteroid production.

End Organ Hypersensitivity (Idiopathic Hirsutism)

There are some patients who present with hirsutism, but ovulate regularly. This category of patients has in the past been labeled idiopathic or familial hirsutism and is more pronounced in certain geographic areas and among certain ethnic groups (especially those from the Mediterranean area). The only satisfactory explanation for this distressing problem is hypersensitivity of the skin’s hair apparatus to normal levels of androgens, probably due to increased 5a-reductase activity. Because of this excessive sensitivity, normal levels of androgen stimulate hair growth. Even in these cases, hirsutism responds to ovarian suppression with a combination oral contraceptive. Suppression of normal female androgen levels to subnormal concentrations diminishes the stimulus to the hair follicles, yielding the same stabilizing results seen in other hirsute women. Spironolactone, finasteride, and finasteride are also effective for this group of patients. Clinical response to pharmacologic treatment correlates with the circulating levels of 3a-AG, supporting the diagnosis of a target tissue (hair follicle) locus for this problem. Thus 5a-reductase inhibition with finasteride may be preferred for this condition (although there have been no comparison studies) because it is directed at the source of the complaint, and there is some evidence that a benefit is sustained after finasteride treatment is discontinued. For this same reason, finasteride may be the treatment of choice for women who have normal circulating androgen levels but suffer from alopecia, a problem due to increased scalp 5a-reductase activity. Although hirsutism due to an endocrine disorder requires control, end organ hypersensitivity is treated only for the purpose of cosmetic improvement. Electrolysis is a useful adjunct in this group of patients.

Summary of Recommendations for the Treatment of Hirsutism

1. The initial treatment of choice for anovulatory women with hirsutism is a low-dose (less than 50 µg estrogen) oral contraceptive.
2. Response is relatively slow, and at least 6 months of treatment are required to demonstrate an impact.
3. When patients do not respond well to oral contraceptives, an antiandrogen should be added, preferably (in order) spironolactone or finasteride.
4. There is no overwhelming evidence that one agent is better than another, and choices should be governed by cost and side effects.
5. The addition of a GnRH agonist should be reserved for patients resistant to initial therapy.
6. Finasteride may be more effective for idiopathic hirsutism and androgenic alopecia.

Limitations and Pitfalls

We have outlined a simple, straightforward approach for the evaluation and management of the hirsute woman; however, as in all of medicine, exceptions occur.

1. Occasionally, testosterone levels may be extremely elevated with anovulation, leading to very heavy hair growth and even masculinization. A testosterone level over 200 ng/dL does not absolutely indicate the presence of a tumor.
2. Enlarged ovaries are not necessary for the clinical syndrome of anovulation and excessive androgen production. On the other hand, the presence of enlarged, polycystic ovaries does not ensure the diagnosis of anovulation and excess ovarian production of androgen. They can be associated with adrenal disease or exogenous androgen ingestion.
3. Laparoscopy and ovarian biopsy are not indicated procedures in the evaluation of hirsutism.
4. The association of elevated testosterone production and hirsutism with normal ovulatory cycles should make the clinician suspicious of an adrenal problem.
5. Suppression of elevated androgens by progestin treatment does not rule out the presence of an ovarian tumor. Functional ovarian tumors are gonadotropin-dependent and responsive.
6. Failure of progestin treatment to suppress hair growth and testosterone levels after 6–12 months raises the suspicion of adrenal disease or a very small ovarian tumor.
7. Androgen levels in postmenopausal women are lower. In this age group, a testosterone level greater than 100 ng/dL is suspicious for a tumor.
149. Carmina E, Lobo RA, Gonadotrophin-releasing hormone agonist therapy for hirsutism is as effective as high dose cyproterone acetate but results in a longer remission, Hum Reprod 12:663, 1997.
The Premenstrual Syndrome

The Premenstrual Syndrome

The simplest definition of the premenstrual syndrome (PMS) is a commonsense one: the cyclic appearance of one or more of a large constellation of symptoms (over 100) just prior to menses, occurring to such a degree that lifestyle or work is affected, followed by a period of time entirely free of symptoms. The most frequently encountered symptoms include the following: abdominal bloating, anxiety or tension, breast tenderness, crying spells, depression, fatigue, lack of energy, unprovoked anger or irritability, difficulty concentrating, thirst and appetite changes, and variable degrees of edema of the extremities — usually occurring in the last 7 to 10 days of the cycle. The exact collection of symptoms in an individual is irrelevant; the diagnosis is made by prospectively and accurately charting the cyclic nature of the symptoms. However, the symptoms are not to be underrated; the various symptoms of premenstrual syndrome have been recounted time and time again in clinicians’ offices in poignant detail, often noting a feeling of being “overwhelmed” or “out of control.”

When women’s daily moods are prospectively charted, a subgroup emerges in which mood changes demonstrate a cyclic pattern with increasing symptoms during the luteal phase and an elimination of symptoms at or soon after menses. Fewer than 50% of women who complain of premenstrual syndrome can be demonstrated to have a pattern of mood changes with a cyclic pattern.

There are two established guidelines for the diagnosis of PMS. The first is from the American Psychiatric Association (APA) and consists of the criteria for what the APA has designated as the premenstrual dysphoric disorder. According to the American Psychiatric Association, this disorder should be differentiated as more severe than PMS; however, most assuredly, there is a broad spectrum of severity, and this differentiation is not helpful or useful. The APA criteria for diagnosis (which correspond to our simple definition above) are as follows:

A. Symptoms are temporally related to the menstrual cycle, beginning during the last week of the luteal phase and remitting after the onset of menses.
B. The diagnosis requires at least 5 of the following, and one of the symptoms must be one of the first 4:
   1. Affective lability; e.g., sudden onset of being sad, tearful, irritable, or angry.
   2. Persistent and marked anger or irritability.
   3. Anxiety or tension.
   4. Depressed mood, feelings of hopelessness.
   5. Decreased interest in usual activities.
   6. Easy fatigability or marked lack of energy.
   7. Subjective sense of difficulty in concentrating.
   8. Changes in appetite, overeating, or food craving.
   9. Hypersomnia or insomnia.
   10. Feelings of being overwhelmed or out of control.
   11. Physical symptoms, such as breast tenderness, headaches, edema, joint or muscle pain, weight gain.
C. The symptoms interfere with work or usual activities or relationships.
D. The symptoms are not an exacerbation of another psychiatric disorder. Thus, PMS is, in part, a diagnosis of exclusion.

The guidelines from the National Institute of Mental Health (NIMH) state that the diagnosis of PMS requires the documentation of at least a 30% increase in severity of symptoms in the 5 days prior to menses compared with the 5 days following menses. Using the NIMH and APA criteria, it is estimated that about 5% of women of reproductive age can be diagnosed with disruptive PMS.

Approximately 40% of women report problems related to their cycles, and about 2–10% report a degree of impact on work or lifestyle. The exact prevalence, however, is difficult to ascertain. The symptoms are variable and difficult to quantify. A further problem that complicates the evaluation of published studies, as well as dealing with individual cases, is that behavior is usually related to menstruation in a retrospective fashion. This is prone to considerable subjective bias. For example, studies in the literature point out that some women do not actually experience problems in relation to menstruation but believe that they do. It is argued, rather convincingly, that men and women in our culture have been conditioned to expect symptoms in a woman’s premenstrual phase and have been taught to expect fluid retention, pain, and emotional reactions. These stereotypic expectations are precisely what are reported when retrospective charting is utilized. Most importantly, carefully constructed studies (prospective with appropriate statistical analyses) show no significant variation associated with the cycle for cognitive, motor, or social behavior.

Is PMS due to an individual pathologic problem or is it due to cultural beliefs, beliefs that lead to the menstrual cycle being associated with a variety of negative reactions, or a combination of both? There is a significant correlation between menstrual symptoms in daughters and mothers, and between sisters, suggesting that these are responses that can be learned. Throughout our recorded history, we find evidence of menstrual taboos. What if our societies and cultures had celebrated menstruation as a time of pleasure (and even public joy) rather than something private (to be hidden) and negative? Would we have PMS today? The answer may lie in the unraveling of the role of our shared beliefs about menstruation in society, rather than the functioning of those beliefs in individuals.

It is generally recognized that R. T. Frank, chief of obstetrics and gynecology at Mt. Sinai Hospital in New York City, first defined premenstrual syndrome in 1931. His
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 unbeatable 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.

Frank went on to summarize 15 cases. He reported that he could obtain relief by withdrawing blood from his patients, and, therefore, theorized that the problem was due to an excess of female sex hormones because of inadequate excretion. Accordingly, he used treatments to enhance excretion, such as calcium lactate, caffeine, and laxatives. For severe cases, he produced ovarian failure by irradiation. S. Leon Israel, in the 1930s, was the first to propose the counterargument that the syndrome was due to defective luteinization resulting in a progesterone deficiency and only a relative hyperestrogenic state. The phrase, “premenstrual syndrome,” was first used in 1953 by Dalton in a report of 84 cases with Greene.

For many years, we were handicapped by a lack of knowledge as to what the premenstrual syndrome really is, how to establish a diagnosis, and how best to treat the condition. In the last decade, however, a degree of understanding has emerged, and treatment guidelines are now available derived from clinical studies.

However studies on premenstrual symptoms that have appropriate controls and statistical treatment find no significant

Etiologies and Treatments

Where scientists have failed to provide proof, practitioners have seldom failed to provide theories. The list of biological theories is impressive:

- Low progesterone levels
- High estrogen levels
- Falling estrogen levels
- Changes in estrogen/progesterone ratios
- Increased aldosterone activity
- Increased renin-angiotensin activity
- Increased adrenal activity
- Endogenous endorphin withdrawal
- Subclinical hypoglycemia
- Central changes in catecholamines
- Response to prostaglandins
- Vitamin deficiencies
- Excess prolactin secretion

Studies prior to 1983 did not incorporate appropriate diagnostic criteria and, therefore, suffer from inaccuracy and heterogeneity. Since 1983, efforts to isolate a...
specific pathophysiologic mechanism have failed to demonstrate differences between women with and without symptoms for all hormone levels throughout the menstrual cycle (including estrogens, progesterone, testosteron, follicle-stimulating hormone, luteinizing hormone, prolactin, and sex hormone-binding globulin) or weight gain and measurements of substances involved in fluid regulation, such as aldosterone. This further includes both the circulating levels as well as the pattern of secretion over the menstrual cycle. In addition, no connection could be demonstrated between PMS and two endogenous metabolites of progesterone, allopregnanolone and pregnanolone. Dynamic electron spectroscopy reveals no abnormalities in the hypothalamic-pituitary axis and its relationships with the adrenal glands, the thyroid gland, and the ovaries. No differences can be detected in magnesium, zinc, vitamin A, vitamin E, thiamin, or vitamin B₆. Some have argued for a greater change in endorphins, proposing that the luteal phase syndrome complex is due to a withdrawal from endogenous opioids (in effect, a autoaddiction and withdrawal), but others have been unable to detect a difference in circulating endorphins in symptomatic patients.

Studies are complicated by high placebo responses. Clinical studies of premenstrual syndrome typically demonstrate a 30–50% response to placebo and, if a specific syndrome? A syndrome must have a specific pathophysiology; specific signs and symptoms can be documented; and a specific treatment.

Is there a specific syndrome? A syndrome must have a specific pathophysiology; specific signs and symptoms can be documented; and a specific treatment.

The experimenter expectancy effect has to be properly controlled. Subjects tend to comply with what they think is the experimenter's hypothesis. This has been experience is a strong statement of the power of the placebo response (in this case, a negative response). In general, the response to ovarian suppression with GnRH agonist-steroid add-back study, patients receiving a placebo instead of the improvement in symptoms. However, the beneficial impact was still considerable.

While medical and surgical oophorectomy has many advocates, especially for women who are a major complaint of bloating; however, appropriate double-blind, placebo-controlled trials have failed to demonstrate a clinical impact (other than on bloatedness) greater than placebo. It has been argued that patients with PMS have a deficiency in fatty acid metabolism, and evening of primrose oil has been advised for therapy. Evening of 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 found to find a difference comparing primrose oil with placebo. The one positive study used retrospective assessment of symptoms, a method known to be inaccurate. Significant improvement has been noted with the use of progesterol in the premenstrual syndrome inhibitors, but it is difficult to know if this is influenced by a positive impact on dysmennorhea.

In the past, there was an enormous amount of publicity given to the use of progesterone treatment by injection or vaginal suppository, long proposed and promoted by Dalton. Four early studies that failed to detect a positive effect of progesterone were criticized for study size and progesterone dosage. A very well-designed study attempted to remove a placebo effect by providing no contact with the investigators or any health care providers during the course of the study; both progesterone and placebo failed to achieve an improvement in symptoms. The criticism of study size and progesterone dose was effectively answered in a randomized placebo-controlled, double-blind, clinical crossover trial of 168 women. Progesterone in doses of 400 mg and 800 mg (doses used by Dalton) did not differ from placebo. A later study by this same group utilizing a very large dose (1200 mg daily) also found no difference between progesterone and placebo treatment.

Only one study has reported beneficial effects with progesterone, a study of only 23 highly motivated women, and the major effect occurred only in the first month of treatment.

Medical and surgical oophorectomy has been described to have dramatic success. A lasting response to surgical hysterecetomy and oophorectomy was reported in women unresponsive to medical therapy. Gonadotropin-releasing hormone (GnRH) agonist treatment can produce hypogonadotropic hypogonadism, in effect, a medical oophorectomy. GnRH agonist treatment has been effective; adding estrogen-progesterin to avoid the side effects of the GnRH agonist diminished somewhat the improvement in symptoms. However, the beneficial impact was still considerable. While medical and surgical oophorectomy is undoubtedly effective, it is impossible to blind such treatment, and the mechanism is, therefore, uncertain. In the GnRH agonist-steroid add-back study, patients receiving a placebo instead of estrogen-progesterin had a return of symptoms (despite continued GnRH agonist treatment), probably in anticipation of a negative reaction to estrogen-progesterin. This experience is a strong statement of the power of the placebo response (in this case, a negative response). In general, the response to ovariian suppression with GnRH agonist treatment indicates that the symptoms of PMS represent an abnormal response to normal hormonal changes.

The only randomized trials, double-blinded and placebo-controlled, which have had consistent, excellent results are those with the antidepressants, fluoxetine, and the one positive study used retrospective assessment of symptoms, a method known to be inaccurate. Significant improvement has been noted with the use of progesterol in the premenstrual synthesis inhibitors, but it is difficult to know if this is influenced by a positive impact on dysmennorhea.

The Ruble study is now a classic. In this study, 44 undergraduates at Princeton University were deliberately deceived about which phase of the menstrual cycle they were experiencing. A bogus electroencephalogram, complete with electrodes attached to the head, was heralded as a new technique capable of predicting the date of menstruation. Subjects were told they were either premenstrual (due in 1–2 days) or intermenstrual (due in 7–10 days). Only those women who were led to believe that their period would begin in 2 days reported significantly higher symptom ratings on pain, water retention, and eating habit changes. This was interpreted as a realization of stereotypic expectations.

The one positive study used retrospective assessment of symptoms, a method known to be inaccurate. Significant improvement has been noted with the use of progesterol in the premenstrual synthesis inhibitors, but it is difficult to know if this is influenced by a positive impact on dysmennorhea.

The Ruble study is now a classic. In this study, 44 undergraduates at Princeton University were deliberately deceived about which phase of the menstrual cycle they were experiencing. A bogus electroencephalogram, complete with electrodes attached to the head, was heralded as a new technique capable of predicting the date of menstruation. Subjects were told they were either premenstrual (due in 1–2 days) or intermenstrual (due in 7–10 days). Only those women who were led to believe that their period would begin in 2 days reported significantly higher symptom ratings on pain, water retention, and eating habit changes. This was interpreted as a realization of stereotypic expectations.

3. Is there a specific syndrome? A syndrome must have a specific pathophysiology; specific signs and symptoms can be documented; and a specific treatment achieves a beneficial response in most patients. One of the basic problems is that we have lumped everything into PMS, including behavioral changes, somatic complaints, and psychological problems. This problem is due to the fact that the techniques of research reflect the way the author of the tool conceptualizes PMS, which, in turn, is based on the background and training of the author.

4. The experimenter expectancy effect has to be properly controlled. Subjects tend to comply with what they think is the experimenter’s hypothesis. This has been studied in regard to PMS, and no significant difference in PMS symptomatology can be demonstrated when the purpose of the study is disguised, and, in addition, the responses can be influenced by positive or negative manipulations. This relates to findings of negative mood changes when subjects are asked to assess their menstrual distress retrospectively.

5. Studies are complicated by high placebo responses. Clinical studies of premenstrual syndrome typically demonstrate a 30–50% response to placebo and, if a positive effect is anticipated by the subjects, up to 80%. Only well-designed, double-blind, placebo-controlled, randomized trials yield reliable data.

The Placebo Response
There is a 3-fold increase in prostaglandin levels in the endometrium from the follicular phase to the luteal phase, with a further increase during menstruation.

Menstruating adolescents were reported to have some dysmenorrhea, and 14% regularly missed school. The severity of dysmenorrhea was directly related to the duration and amount of menstrual flow. In a longitudinal American study, 8% missed school or work at every menses, and 38.2% regularly used medical treatment.

Dysmenorrhea is pain with menstruation, usually cramping in nature and centered in the lower abdomen. Studies on the prevalence of dysmenorrhea are few. In a placebo-controlled, randomized trial to be associated with a 48% reduction in symptom scores (compared with a 30% reduction in the placebo-treated group).

Treatment of Premenstrual Syndrome

The first step is to be convinced (both patient and clinician) that the problem is cyclic. The only instrument of diagnosis available at the present time is the menstrual calendar. There is no single calendar that has emerged as superior and acceptable to all; however, several are available in the literature.

We offer our perspective on this syndrome, suggesting that it is not a single disorder, but rather a collection of different problems. We believe that PMS has a basic psychobiologic origin, tied to the menstrual cycle, primarily biologically, but with a psychosociologic overlay. This can be a learned response or it can be a response in vulnerable individuals triggered by normal neuroendocrine and hormonal changes. The hormonal changes of the menstrual cycle are not an etiologic factor, but they can operate to produce in susceptible women mood changes or a destabilization of mood, and specifically involving the serotonergic system.

Sometimes, patients present to the clinician totally focused on complaints that occur premenstrually. With exploration of lifestyle, relationships, and interactions, the focus on a premenstrual syndrome can be shifted to the underlying issues that are producing conflict and lack of control. Helping a patient to come to grips with the subtle nature of this problem, the fundamental psychologic response involved, and the need to take charge of one’s life represent the type of broad involvement required of a clinician. Without this type of broad involvement, only a short-term response can be achieved with little hope for long-term success.

Any changes that allow individuals to exert greater control over their lives will produce a positive impact. It is for this reason that lifestyle changes are effective in the treatment of PMS. Changes in diet, changes in exercise, changes in work or recreation — all are examples of exerting control over life rather than having life's circumstances control the individual.

If the practitioner is convinced of the cyclic nature of a problem (by a prospective record of at least 3 months duration), try to isolate the specific symptoms and treat with a specific therapy. If fluid retention is perceived by the patient as a principal problem, offer diuretic therapy with spironolactone. If dysmenorrhea is a component of the symptom complex, try one of the inhibitors of prostaglandin synthetase or oral contraceptives. Calcium supplementation (1200 mg daily) was observed in a placebo-controlled, randomized trial to be associated with a 48% reduction in symptom scores (compared with a 30% reduction in the placebo-treated group).

These are safe and relatively inexpensive approaches that deserve initial consideration.

A failure to identify a specific disorder with a specific mechanism suggests that premenstrual syndrome represents a variety of psychological manifestations triggered by normal, cyclic hormonal changes. This process can be either physiologic in nature or psychosocial and deeply rooted in our cultural history. For that reason, it makes some sense to completely eliminate endogenous sex steroid variability. This can be achieved with daily oral contraceptives, or medroxyprogesterone acetate, 10–30 mg daily, or depot-medroxyprogesterone acetate, 150 mg every 3 months. On occasion, we have induced beneficial and gratifying results in patients with cyclical mood and behavioral swings. But in view of the vague and subjective nature of this syndrome, any such empiric therapeutic treatment must be pursued in a fully informed fashion. If a patient is willing to undergo an empiric trial, we are willing. In doing so, however, neither partner in this contract should be deceived; we must remember that the placebo response may be the underlying basis for any positive response. But keep in mind that the placebo response is another example of an individual exerting control over life. In this case, it represents the subtle effort of the body at a subconscious level to exert self-healing.

The 工" treatment of PMS begins with patient confidence in the clinician and extends through to the full functioning of the patient's own healing system. Interaction with the clinician provides a better understanding of what's going on (at least some elimination of unfounded fears) and provides some hope. Many of the treatment modalities for PMS, if not all, provide a woman with a greater sense of control over life; thus, minimal interaction, such as focusing on diet or lifestyle, can yield a positive result. In the process of making detailed, prospective observations of one's own life, a patient can experience an increase in self-control, which is in itself a therapeutic process.

Leon Eisenberg has written the following insightful and helpful thoughts on the placebo:

"The strange sounding word, placebo, comes from the Latin verb meaning "I shall please." Clinicians and patients have been educated to observe a prescription ritual. Most people seem to feel that their complaints are not taken seriously unless they are in possession of a prescription. But the placebo is not so much a pill as a process."

The process begins with patient confidence in the clinician and extends through to the full functioning of the patient's own healing system. Interaction with the clinician provides a better understanding of what's going on (at least some elimination of unfounded fears) and provides some hope. Many of the treatment modalities for PMS, if not all, provide a woman with a greater sense of control over life; thus, minimal interaction, such as focusing on diet or lifestyle, can yield a positive result. In the process of making detailed, prospective observations of one's own life, a patient can experience an increase in self-control, which is in itself a therapeutic process.

"So empathically 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 component 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."

Until PMS is better understood, the placebo response will continue to play an important role in therapy. There is a psychosocial subjective component of medicine that makes the placebo process a legitimate part of every patient–clinician interaction.

Dysmenorrhea

Dysmenorrhea is pain with menstruation, usually cramping in nature and centered in the lower abdomen. Studies on the prevalence of dysmenorrhea are few. In a random sample of 19-year-old women in Gothenburg, Sweden, 72% reported dysmenorrhea, 15% had to limit their daily activity and the severity was unimproved by analgesics, 8% missed school or work at every menses, and 38.2% regularly used medical treatment. Oral contraceptive use and previous vaginal deliveries were associated with less dysmenorrhea. The severity of dysmenorrhea was directly related to the duration and amount of menstrual flow. In a longitudinal American study of 17–19-year-old university students, 13% reported severe pain in more than half of their menstrual periods, and 42% indicated interference with activity at least once. A survey in Turkey reported that 25.6% of adolescents missed school because of dysmenorrhea. In another older but good prevalence study, 45% of surveyed women had moderate or severe dysmenorrhea. Most adolescents experience dysmenorrhea in the first 3 years after menarche. In the U.S., about 60% of menstruating adolescents were reported to have some dysmenorrhea, and 14% regularly missed school. Menstrual problems, including dysmenorrhea, are more common in individuals who smoke. Dysmenorrhea improves in most women after a full-term pregnancy.

Primary dysmenorrhea, a condition associated with ovulatory cycles, is due to myometrial contractions induced by prostaglandins originating in secretory endometrium, while secondary dysmenorrhea is associated with a variety of pathological conditions. Secondary dysmenorrhea is more common in women who smoke. Dysmenorrhea improves in most women after a full-term pregnancy.
Women with primary dysmenorrhea have greater endometrial production of prostaglandins than asymptomatic women. Most of the release of prostaglandins during menstruation occurs during the first 48 hours, which coincides with the greatest intensity of the symptoms. Prostaglandin F₂α (PGF₂α) is the agent responsible for dysmenorrhea. It always stimulates uterine contractions, whereas the E prostaglandins inhibit contractions in the nonpregnant uterus. Uterine muscle from both normal and dysmenorrheic women is sensitive to PGF₂α, but the amount of PGF₂α produced is the major differentiating factor.

The clinical benefit derived from the pharmacologic use of inhibitors of prostaglandin synthesis depends on a significant decrease in prostaglandin production in the endometrium, as well as shortened and diminished menstrual flow. An additional role may be attributed to decreased prostaglandins from the platelets participating in the clotting of menstrual blood. The explanation for the benefit seen with oral contraceptives is decreased prostaglandin synthesis associated with the atrophic decidualized endometrium. Oral contraception is a good choice for therapy, combining contraception with a beneficial impact on dysmenorrhea, menstrual flow, and menstrual irregularity.

In women who do not desire hormonal contraception, the best therapy is one of the agents that inhibit prostaglandin synthesis. There are several families of nonsteroidal anti-inflammatory agents. The acetic acid group is associated with more side effects, and these agents are not the drugs of choice for dysmenorrhea; indomethacin belongs to this group. The propionic acid derivatives (ibuprofen, naproxen, ketoprofen) and the fenamates (mefenamic acid, meclofenamate, flufenamic acid) are very effective for the treatment of dysmenorrhea.

The findings in 51 clinical trials of prostaglandin synthetase inhibitors indicated that the fenamates are more effective, but not substantially, for relieving pain. The fenamates, in addition to inhibiting prostaglandin synthesis, also have an antagonistic action, competing for prostaglandin-binding sites. Side effects associated with these agents are minimal, but can include blurred vision, headaches, dizziness, and gastrointestinal discomfort. The latter can be reduced by taking the medication with milk or food. All of these agents are more potent than aspirin, because the uterus is relatively insensitive to aspirin. The major contraindications to the use of these agents include gastrointestinal ulcers and hypersensitivity to aspirin and similar agents.

Approximately 80% of dysmenorrheic women are relieved by prostaglandin inhibitors. Improvement is noted in the symptoms associated with menses, specifically cramping, backache, nausea, vomiting, dizziness, leg pain, insomnia, and headache. A trial of up to 6 months is warranted, with necessary changes in dosage and inhibitors, before abandoning this therapy. Initially, it was believed that better relief was achieved if treatment was started 2–3 days before menses in order to lower the tissue level of prostaglandins before breakdown of the endometrium. Fortunately, studies have indicated that treatment is just as effective if begun at the sign of first bleeding, thus decreasing the possibility of taking one of these agents early in pregnancy. Another benefit of prostaglandin inhibition is a reduction in the amount of blood lost with periods. Indeed, the agents may be used to treat idiopathic menorrhagia, or the excess flow associated with an intrauterine device (IUD). Most women do not need to take the medication more than 2–3 days.

If dysmenorrhea is not relieved by one of the nonsteroidal, anti-inflammatory analgesics, laparoscopy should be seriously considered to determine the cause of the symptoms. Conditions associated with dysmenorrhea include müllerian duct anomalies, endometriosis, and pelvic inflammatory disease. We should especially be aware that endometriosis occurs in adolescents; dysmenorrhea caused by endometriosis in adolescents usually begins 3 or more years after menarche.

Menstrual Headache

Headaches are very common, but it is rare when the cause of the headache is a serious problem. Most headaches are due to vasodilatation, muscle contraction, or psychologic stress. Menstrual headaches include all headaches related in temporal fashion to menses, beginning before or during menstrual flow. For many women with premenstrual syndrome, headache is part of the constellation of PMS symptoms. Here, we are considering the occurrence of headache as a single, solitary symptom associated with menses.

Migraine headaches have a peak incidence of first occurrence at age 15–19; they are most prevalent in women in their late 30s to early 40s, and rare after menopause. An association with menses is observed by 60% of women with migraine headaches. In 7–14% of women with migraine, headaches occur exclusively with menses. Because menstrual migraine improves in two-thirds of migraineurs with pregnancy, this type of migraine seems to be due to falling levels of estrogen and progesterone, triggering a host of responses such as release of prostaglandins and changes in neurotransmitters.

Classic migraine is associated with a visual aura, including bright lights, zigzag lights, or flashes and sparkles. Common migraine is not associated with neurologic auras. Complicated migraine refers to headache associated with dramatic focal neurologic features that are transient; e.g., blindness, unilateral paresthesias, or paresis. With increasing age, the focal neurologic phenomena can occur without headache.

Vascular Headaches

Acute and throbbing headaches are due to abnormal vasodilatation. The vasodilatation associated with migraine headaches is believed to follow a period of vasoconstriction. Migraine headaches are usually, but not always, preceded by prodromal symptoms (which may reflect the period of vasoconstriction). Significant vascular headaches can be precipitated by stress, alcohol, or tyramine and tryptophan rich foods (red wine, chocolate, ripe cheeses). Vascular headaches can accompany other problems, such as systemic viral infections, fever, or hypertension. Common migraine headaches are known as “migraine without aura.” Classic migraine is referred to as “migraine with aura.” Migraine headaches associated with menstruation are typically migraine without aura.

Tension Headaches

The common tension headache is due to prolonged and excessive muscle contraction. The pain is dull, steady, bilateral, and worsens throughout the day. The headaches frequently occur with worry or emotional stress and commonly last for hours or a couple of days; however, everyday minor hassles are a more important factor in the pathogenesis of tension headaches than major stressful events.

Secondary Headaches

This type of headache is due to underlying organic disease. The pain is usually due to pressure or pulling of structures. Headaches associated with brain tumors are usually accompanied by neurologic abnormalities. Other causes are brain abscesses, subdural hematomas, hypertension, drug use, and concussions. The main cause of inflammatory headaches is meningitis.

Evaluation

The acute onset of severe headache pain deserves attention. The following signs suggest the presence of a serious problem: neck stiffness, altered mental status, focal neurologic abnormalities, visual impairment, and fever. Any patient with meningeal signs requires hospitalization. Keep carbon monoxide exposure and drug withdrawal in mind as etiologic agents.

Chronic headaches should be characterized according to location, quality, and course over time. Head trauma in the past is an important piece of information, raising the suspicion of a subdural hematoma. When the headache is cyclic, with periodic complete resolution, one can comfortably ascribe the headache to a vascular origin. Tension headaches are either variable or relatively constant without relentless progression. Any recurrent or chronic headache that gets worse with time deserves a neurologic evaluation.

Management

Menstrually related migraines are more refractory to the battery of therapy used by neurologists, although oral and subcutaneous sumatriptan has been effective. The nonsteroidal anti-inflammatory agents are also effective, particularly when headaches are associated with dysmenorrhea.
Early studies of menstrual migraine indicated that administration of estrogen could delay the onset of migraine even if menses were not delayed. Progestrone administration delayed menses, but not the onset of headache. Others have claimed effective treatment of menstrual migraine with maintenance of estrogen levels (there is reason to believe that a relatively high estrogen level is necessary; e.g., the use of the 100 μg transdermal patch). Still others have reported success with tamoxifen or danazol treatment. Unfortunately this field suffers from a lack of well-designed, double-blind, placebo-controlled studies, and we must make our judgments based on experience.

We have had personal success (anecdotal to be sure) alleviating headaches by eliminating the menstrual cycle, either with the use of daily oral contraceptives (not the progestin-only minipill) or the daily administration of a progestational agent (such as 10 mg medroxyprogesterone acetate). Some menstrual migraineurs have extremely gratifying responses.

If menstrual headaches are a reaction to cyclic changes in circulating levels of the sex steroids, it makes sense to avoid cyclically and maintain a relatively steady state with daily administration of exogenous hormones. A more expensive approach, but one with documented success, is the utilization of a long-acting GnRH agonist combined with estrogen-progestin add-back therapy. Another option is to use an estrogen transdermal application during the menstrual time period.

The elimination of cyclicity can be applied to postmenopausal women who experience exacerbation or onset of headaches on a sequential hormone regimen. The maintenance of daily, relatively constant hormone levels with the daily, continuous program of combined estrogen-progestin has been effective in our experience.

The run-of-the-mill headache is treated with mild analgesics such as aspirin, acetaminophen, or the nonsteroidal anti-inflammatory agents. A problem of severe headaches on oral contraception requires an immediate response. The conservative reaction is to discontinue the oral contraceptives. On the other hand, the headache pattern may not be easily explained by the normal menstrual or cyclic changes and requires a more diagnostic approach. The mechanism is not known, but obvious effects of menstrual hormonal changes have been suggested, such as prostaglandin release, changes in the immune system, or a direct impact of declining estrogen and progesterone on bronchial smooth muscle. The administration of estrogen (estradiol 2 mg orally daily) improves symptoms and measurements of pulmonary function. In another report, the administration of intramuscular progesterone ameliorated premenstrual asthma. Once again, the logic is appealing that the elimination of menstrual periods (with the daily administration of oral contraceptives or intramuscular depot-medroxyprogesterone acetate, or GnRH agonist treatment with estrogen-progestin add-back) can be used for this purpose.

Catamenial Seizures

Catamenial epilepsy in ancient times was attributed to the moon, giving rise to the word, “lunatic.” Epileptic seizures increase in frequency during menstruation and decrease during the luteal phase. Exacerbation of seizure activity with menses occurs in 50% of epileptic women. In addition, seizure frequency increases at the time of the midcycle peak in estrogen and during anovulatory cycles. In animal experiments, estrogen increases seizure activity, and progesterone is antiepileptic. These observations suggest an antiepileptic effect of progesterone.

Progesteronal hormones are known to have a sedative effect on the central nervous system. This pharmacologic effect combined with the observations indicating increased seizure activity at times when circulating levels of progesterone are low indicated that treatment with a progestin would have a beneficial impact on seizures. Treatment with orally administered progesterone during the luteal phase has been reported to decrease seizure activity. Antiepileptic drugs enhance hepatic metabolic activity, and, therefore, doses must be relatively high. Oral medroxyprogesterone acetate is relatively ineffective, probably because it is difficult to achieve high blood levels. Intramuscular injections of depot-medroxyprogesterone acetate can improve seizure control. Depot-medroxyprogesterone acetate, 150 mg im every 1–2 months, can decrease seizure frequency by approximately 50%. In a case report of an 8-year-old girl, 150 mg administered every 2 weeks abolished seizure activity. Intravenous progesterone (producing luteal phase levels) can produce a significant decrease in spike frequency.

In women with a seizure pattern consistently linked to menstruation, elimination or suppression of menstruation can bring impressive relief. The administration of daily oral contraceptives, or intramuscular depot-medroxyprogesterone acetate, or GnRH agonist treatment with estrogen-progestin add-back can be used for this purpose.

Premenstrual Asthma

Approximately 30–40% of women with asthma have an increase in symptoms associated with menstruation. Even asthmatics not aware of a link to menstruation demonstrate a menstrual worsening of pulmonary function. The mechanism is not known, but obvious effects of menstrual hormonal changes have been suggested, such as prostaglandin release, changes in the immune system, or a direct impact of declining estrogen and progesterone on bronchial smooth muscle. The administration of estrogen (estradiol 2 mg orally daily) improves symptoms and measurements of pulmonary function. In another report, the administration of intramuscular progesterone ameliorated premenstrual asthma.

In women with a seizure pattern consistently linked to menstruation, elimination or suppression of menstruation can bring impressive relief. The administration of daily oral contraceptives, or intramuscular depot-medroxyprogesterone acetate, or GnRH agonist treatment with estrogen-progestin add-back can be used for this purpose.

Premenstrual Pneumothorax

Very rarely, recurrent pneumothoraces, often related to pulmonary endometriosis, occur at the time of menses. Effective medical treatment requires the long-term suppression of menstruation. The two best therapeutic choices are again the daily administration of oral contraceptives or treatment with a long-acting GnRH agonist supplemented with a daily estrogen-progestin combination.

Chapter References

Dysfunctional uterine bleeding is defined as a variety of bleeding manifestations of anovulatory cycles (in the absence of pathology or medical illness). It can be confidently managed without surgical intervention by therapeutic regimens founded on sound physiologic principles. Our formulation is based on knowledge of how the postovulatory menstrual function is naturally controlled, and uses pharmacologic application of sex steroids to reverse the abnormal tissue factors that lead to the excessive and prolonged flow typical of anovulatory cycles.

Three major categories of dysfunctional endometrial bleeding are dealt with:

1. Estrogen withdrawal bleeding,
2. Estrogen breakthrough bleeding, and
3. Progestin breakthrough bleeding.

In each instance, the manner in which the endometrium deviates from the norm is characterized, and specific sex steroid therapy is recommended to counter the difficulties each situation presents.

This mode of clinical management has been in regular use for many years, and failure to control vaginal bleeding with this therapy, despite appropriate application and utilization, excludes the diagnosis of dysfunctional uterine bleeding. If this occurs, attention is directed to a pathologic entity within the reproductive tract as the cause of abnormal bleeding.

Heavy but regular menstrual bleeding can be encountered in ovulating women. In the absence of a specific pathologic cause, it is presumed that this reflects subtle disturbances in the endometrial tissue mechanisms discussed in Chapter 4. In essentially all cases, evaluation and treatment are identical to the approach detailed in this chapter.

### 15 Normal Withdrawal (Menstrual) Bleeding

Of all the types of hormonal-endometrial relationships, the most stable endometrium and the most reproducible menstrual function in terms of quantity and duration occurs with postovulatory estrogen-progesterone withdrawal bleeding. It is so controlling that many women over the years come to expect a certain characteristic flow pattern. Any slight deviations, such as plus or minus 1 day in duration or minor deviation from expected napkin or tampon utilization, are causes for major concern in the patient. So ingrained is the expected flow that considerable clinician reassurance may be required in some instances of minor variability. Although variability of menstrual cycles is a common feature during teenage years and the perimenopausal transition, the characteristics of menstrual bleeding do not undergo appreciable change during the reproductive years.1

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 40.2 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. 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.3 Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.4 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.5

The usual duration of flow is 4–6 days, but many women flow as little as 2 days, and as much as 7 days. The normal volume of menstrual blood loss is 30 mL.6 Greater than 80 mL is abnormal. Most of the blood loss occurs during the first 3 days of a period, so excessive flow may exist without prolongation of flow.7

### Menstrual Period Characteristics
Although the postovulatory phase averages 14 days, greater variability in the proliferative phase produces a distribution in the duration of a menstrual cycle. Based on the normal experience, menstrual bleeding more often than every 24 days or less often than every 35 days deserves evaluation. Flow that lasts 7 or more days also deserves evaluation. A flow that totals more than 80 mL per month usually leads to anemia and should be treated. In general, however, an effort to quantitate menstrual blood loss is not necessary because evaluation and treatment are responses to a patient's own perceptions regarding duration, amount, and timing of her menstrual bleeding. Although the correlation between a patient's perceptions and actual menstrual blood loss is poor, an individual patient's anxiety and concern deserve consideration and evaluation. Midcycle bleeding can be a consequence of the preovulatory fall in estrogen; however, intermenstrual bleeding is often due to pathology.

The histologic changes in the endometrium of a normal, ovulatory menstrual cycle have been detailed in Chapter 4. An understanding of these events provides the foundation and framework to understand abnormal endometrial bleeding. There are three reasons for the self-limited character of estrogen-progesterone withdrawal bleeding.

1. It is a universal endometrial event. Because the onset and conclusion of menses are related to a precise sequence of hormonal events, menstrual changes occur almost simultaneously in all segments of the endometrium.

2. The endometrial tissue that has responded to an appropriate sequence of estrogen and progesterone is structurally stable, and random breakdown of tissue due to fragility is avoided. The events leading to ischemic disintegration of the endometrium are orderly and progressive, being related to rhythmic waves of vasoconstriction of increasing duration.

3. Inherent in the events that start menstrual function following estrogen-progesterone are the factors involved in stopping menstrual flow. Just as waves of vasoconstriction initiate the ischemic events, prolonged vasoconstriction abetted by the stasis associated with endometrial collapse enables clotting factors to seal off the exposed bleeding sites. Additional and significant effects are obtained by resumed estrogen activity.

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 following vascular sequence has been constructed from direct observations of rhesus endometrium. With shrinkage of height, blood flow within the spiral vessels diminishes, venous drainage is decreased, and vasoconstriction ensues. Thereafter, the spiral arterioles undergo rhythmic vasoconstriction and relaxation. Each successive spasm is more prolonged and profound, leading eventually to endometrial thinning. Within the 24 hours immediately preceding menstruation, these reactions lead to endometrial ischemia and stasis. While cells migrate through capillary walls, at first remaining adjacent to vessels, but then extend into the interendothelial space. During aging, a further escape into the interstitial space occurs. Thrombin-platelet plugs form intravascularly upstream at the shedding surface, limiting blood loss. Increased blood loss is a consequence of reduced platelet numbers in superficial vessels. The prostaglandin content (PGF₂α and PGE₂) 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 PGF₂α from glandular cells and the potent vasoconstrictor, endothelin-1, derived from stromal decidual cells.

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 constituents, 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.

Endometrial tissue breakdown involves a family of enzymes, matrix metalloproteinases, that degrade components (including collagens, gelatins, fibronectin, and laminin) of the extracellular matrix and basement membrane. The metalloproteinases include collagens that degrade interstitial and basement membrane collagens, gelatins that further degrade collagens, and stromelysins that degrade fibronectin, laminin, and glycoproteins. The expression of metalloproteinases in human endometrium follows a pattern correlated with the menstrual cycle, including 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. Progestosterone withdrawal from endometrial cells induces matrix metalloproteinase secretion that is followed by the breakdown of cellular membranes and the dissolution of extracellular matrix. 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 TGF-β. In a nonpregnant cycle, metalloproteinase expression is suppressed after menstruation, presumably by increasing estrogen levels. Metalloproteinase activity is restrained by specific tissue inhibitors, designated as TIMP. Thus progesterone withdrawal can lead to endometrial breakdown through a mechanism that is independent of vascular events (specifically ischemia), a mechanism that involves cytokines.

Eventually considerable leakage occurs as a result of diapedesis, and, finally, interstitial hemorrhage occurs due to breaks in superficial arterioles and capillaries. As ischemia and weakening progress, the continuous binding fragment is maintained, 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 enzymatic action, 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. PAI-1, in particular, exerts an important restraining action on fibrinolysis and proteolytic activity.

With further tissue disorganization, the endometrium shrinks even more, and coiled arterioles are buckled. Additional ischemic breakdown ensues with necrosis of cells and defects in vessels adding to the menstrual effluxuum. 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 inexcogitably extends throughout the uterus. In the end, the typical deflated shallow dense menstrual endometrium results. Within 13 hours, the endometrial height shrinks from 4 mm to 1.25 mm. Menstrual flow stops as a result of the combined effects of prolonged vasoconstriction, 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 provides the generation of fibrin, the activation of platelets and clotting factors, and angiogenesis.

Platelets and fibrin play a direct part in the hemostasis achieved in a bleeding menstrual endometrium. Deficiencies in these constituents cause the increased blood loss seen with von Willebrand’s disease and in thrombocytopenia. The blood loss at menses in albinogen's disease indicates the importance of fibrin-generating and fibrinolytic factors in the menstrual process. Intravascular thrombi are observed in the functional layers and are localized to the shedding surface of the tissue. These are known as ‘plugging’ plugs because blood may flow past these only partially occlusive barriers. Therefore, thrombin continues to develop within the menstrual blood, accounting for the platelets and large amounts of fibrin found in this effluent. Fibrinolysis occurs in the endometrial tissue, limiting fibrin deposition in the proximal, still unshaped layer. Despite large holes in vessel walls, with blood exposed to collagen surfaces, no occlusive surface thrombus is formed. After early tissue dependence on thrombin plugs to restrain blood loss, later generalized vasoconstrictive hemostasis without thrombin plugs occur. The healing endometrium is pale, collapsed, and disordered, but no thrombo and no fibrin deposits are seen.

Lockwood assigns a key role to decidual cells in both the process of endometrial bleeding (menstruation) and the process of endometrial hemostasis (implantation and placentation). Implantation requires endometrial hemostasis and the maternal uterus requires resistance to invasion. Inhibition of endometrial hemorrage 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), increased levels of plasminogen activator inhibitor-1 and
tissue factor. Withdrawal of estrogen and progesterone support, however, leads to changes in the opposite directions, consistent with endometrial breakdown.

Traditional Definitions

<table>
<thead>
<tr>
<th>Oligomenorrhea: Intervals greater than 35 days.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymenorrhea: Intervals less than 24 days.</td>
</tr>
<tr>
<td>Menorrhagia: Regular normal intervals, excessive flow and duration.</td>
</tr>
<tr>
<td>Metrorrhagia: Irregular intervals, excessive flow and duration.</td>
</tr>
</tbody>
</table>

Endometrial Responses to Steroid Hormones: Physiologic and Pharmacologic

Obviously, estrogen and progesterone withdrawal is not the only type of endometrial bleeding provoked by the presence of sex steroids and their effects on the endometrium. There are clinical examples for estrogen withdrawal bleeding and estrogen breakthrough bleeding, as well as for progesterone withdrawal and breakthrough bleeding.

Estrogen Withdrawal Bleeding

This category of uterine bleeding can occur after bilateral oophorectomy, radiation of mature follicles, or administration of estrogen to a castrate and then discontinuation of therapy. Similarly, the bleeding that occurs postcastration can be delayed by concomitant estrogen therapy. Flow will occur on discontinuation of exogenous estrogen. Midcycle bleeding can occur secondary to the decrease in estrogen that immediately precedes ovulation.

Estrogen Breakthrough Bleeding

Here a semiquantitative relationship exists between the amount of estrogen stimulating the endometrium and the type of bleeding that can ensue. Relatively low doses of estrogen yield intermittent spotting that may be prolonged, but is generally light in quantity of flow. On the other hand, high levels of estrogen and sustained availability lead to prolonged periods of amenorrhea followed by acute, often profuse bleeds with excessive loss of blood.

Progesterone Withdrawal Bleeding

Removal of the corpus luteum will lead to endometrial desquamation. Pharmacologically, a similar event can be achieved by administration and discontinuation of progesterone or a nonestrogenic synthetic progestin. Progesterone withdrawal bleeding occurs only if the endometrium is initially proliferated by endogenous or exogenous estrogen. If estrogen therapy is continued as progesterone is withdrawn, the progesterone withdrawal bleeding still occurs. Only if estrogen levels are increased 10–20-fold will progesterone withdrawal bleeding be delayed.

Progesterone Breakthrough Bleeding

Progesterone breakthrough bleeding occurs only in the presence of an unfavorably high ratio of progesterone to estrogen. In the absence of sufficient estrogen, continuous progesterone therapy will yield intermittent bleeding of variable duration, similar to low-dose estrogen breakthrough bleeding noted above. This is the type of bleeding associated with the long-acting progestin-only contraceptive methods, Norplant and Depo-Provera.

Suggestions for Why Anovulatory Bleeding Is Excessive

Most instances of anovulatory bleeding are examples of estrogen withdrawal or estrogen breakthrough bleeding. The heaviest bleeding is secondary to high sustained levels of estrogen associated with polycystic ovaries, obesity, immaturity of the hypothalamic-pituitary-ovarian axis as in postpubertal teenagers, and late anovulation, usually involving women in their late 30s and 40s. In the absence of growth-limiting progesterone and periodic desquamation, the endometrium attains an abnormal height without concomitant structural support. The tissue increasingly displays intense vascularity, back-to-back glandularity, but without an intervening stromal support matrix. This tissue is fragile and will suffer spontaneous superficial breakage and bleeding. As one site heals, another, and yet another new site of breakdown will appear. The typical clinical picture is that of a pale frightened teenager who has bled for weeks. Also frequently encountered is the older woman with prolonged bleeding who is deeply concerned over this experience as a manifestation of cancer.

In these instances, the usual endometrial control mechanisms are missing. This bleeding is not a universal event, but, rather, it involves random portions of the endometrium at variable times and in asynchronous sequences. The fragility of the vascular adenomatous hyperplastic tissue is responsible for this experience, in part because of excessive growth, but mostly because of irregular stimulation in which the structural rigidity of a well-developed stratum or stratum compactum does not occur. Finally, the flow is prolonged and excessive not only because there is a large quantity of tissue available for bleeding, but more importantly, because there is no organized, abrupt, random, breakdown of tissue with consequent opening of multiple vascular channels. There is no vasoconstrictive rhythmicity, no tight coiling of spiral vessels, no orderly collapse to induce stasis. The anovulatory tissue can only rely on the “healing” effects of endogenous estrogen to stop local bleeds. However, this is a vicious cycle in that this healing is only temporary. As quickly as it rebuilds, tissue fragility and breakdown recur at other endometrial sites.

Alternate Hypothesis

Another explanation for the control of postovulatory endometrial bleeding and regeneration has been presented. Based on light and scanning electron microscopy of hysterectomy specimens, this thesis favors nonhormone-related regeneration of surface epithelium from basal glands and cornual area residual tissue with restoration of the continuous binding membrane as the critical events in cessation of blood flow. Endometrial regeneration is viewed as a response to tissue loss, not hormonal changes. By this account, estrogen withdrawal or breakthrough bleeding is uncontrolled because there is insufficient stimulus (loss of tissue) for binding surface restoration to occur. Furthermore, curettage is effective in this condition by reaching sufficient basal glandular denudation (as is seen also in combined estrogen and progestin withdrawal), which stimulates regeneration of surface integrity and, thus, controls blood flow.

Additional studies are needed to clarify the difference of opinion concerning the pathophysiology of dysfunctional uterine bleeding. Our therapeutic approach favored in this book uses hormonal control of endometrial events and rarely finds it necessary to resort to surgery.

Differential Diagnosis

Dysfunctional uterine bleeding is a diagnosis made by exclusion. A very common cause of abnormal uterine bleeding is pregnancy and pregnancy-related problems such as ectopic pregnancy or spontaneous miscarriage. This category of problems should always receive diagnostic consideration because the most common cause of an abrupt departure from a normal menstrual pattern is pregnancy or a complication of pregnancy. Patients may be using medications unknowingly with an impact on the endometrium. For example, the use of ginseng, an herbal root, has been associated with estrogenic activity and abnormal bleeding. Pathology of the menstrual outflow tract includes cancers of the cervix and endometrium, endometrial polyps, leiomyomata uteri, and infections. Although uterine bleeding is a common problem with various contraceptive methods and postmenopausal hormonal therapy, the clinician should always be convinced no pathology is present. Abnormal menstrual cycles are occasionally the first sign of either hypothyroidism or hyperthyroidism. Irregular, serious bleeding is often associated with severe organ disease, such as renal failure and liver failure. Finally, careful examination is worthwhile to discover genital injury or a foreign object.

One should keep in mind that as many as 20% of adolescents with dysfunctional uterine bleeding will have a coagulation defect, although the most common cause is anovulation. In 150 women aged 15–50 in Helsinki who were referred for the evaluation of menorrhagia, an inherited bleeding disorder was diagnosed in 26 of the patients; the most common diagnoses were von Willebrand’s disease (13%) and factor XI deficiency (4%). Bleeding secondary to a blood dyscrasia is usually a heavy flow with regular, cyclic menses (menorrhagia), and this same pattern can be seen in patients being treated with anticoagulants. Bleeding disorders are usually associated with menorrhagia since menarche and a history of bleeding with surgery or trauma. Menorrhagia may be the only sign of an inherited bleeding disorder.

The effects of tubal ligation are still not certain. The first well-controlled studies of this issue demonstrated no change in menstrual patterns, volume, or pain.
Subsequently, these same authors reported an increase in dysmenorrhea and changes in menstrual bleeding. However, these authors 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, but not by others. 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. 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. These discordant reports do not make patient counseling about the long-term effects of tubal sterilization an easy task. It is possible that extensive electrocoagulation of the fallopian tubes can change ovarian steroid production. Perhaps this is why menstrual changes were detected with longer (4 years) follow-up, while no changes have been noted with the use of rings or clips.

However, attempts to relate poststerilization menstrual changes with extent of tissue destruction fail to find a correlation, and an increase in hospitalization for menstrual disorders after unipolar cautery cannot be documented. Still another long-term follow-up study (3–4.5 years) failed to document any significant changes in menstrual cycles. This inconsistency can reflect differences in sterilization techniques, as well as the fact that a surgical solution is more likely to be chosen if continuing fertility is no longer an issue. The best answer for now is that some women experience menstrual changes, but most do not.

Laboratory tests that can be helpful (but not always necessary) are coagulation studies (prothrombin time, partial thromboplastin time, platelet count, bleeding time, and the ristocetin cofactor assay for the von Willebrand factor), complete blood count, quantitative human chorionic gonadotropin (HCG), prolactin, thyroid function tests, liver function tests, and appropriate cervical cultures.

Treatment Program for Anovulatory Bleeding

The immediate objective of medical therapy in anovulatory bleeding is to relieve the natural controlling influences missing in this tissue: universal, synchronous endometrial events, structural stability, and vasomotor rhythmicity.

Progestin Therapy

Most women will, at sometime during their reproductive years, either fail to ovulate or not sustain adequate corpus luteum function or duration. But this occurs with increased frequency in adolescence and in the decade prior to menopause. The usual clinical presentation is oligomenorrhea with bouts of heavy bleeding. Women correctly seek medical advice promptly because these menstrual aberrations suggest unplanned pregnancy or uterine pathology. Under most circumstances, progestin therapy will suffice to control the abnormality once uterine pathology is ruled out.

Progestrone and progestins are powerful antiestrogens when given in pharmacologic doses. Progestins stimulate 17b-hydroxysteroid dehydrogenase and sulfotransferase activity, which convert estradiol to estrone sulfate (which is rapidly excreted from the cell). Progestins also diminish estrogen effects on target cells by inhibiting the augmentation of estrogen receptors that ordinarily accompanies estrogen action (receptor replenishment inhibition). In addition, progestins suppress estrogen-mediated transcription of oncogenes. These influences account for the antimitotic, antigrowth impact of progestins on the endometrium (prevention and reversal of hyperplasia, limitation of growth postovulation, and the marked atrophy during pregnancy or in response to combined oral contraceptives).

In the treatment of oligomenorrhea, orderly limited withdrawal bleeding can be accomplished by administration of a progestin such as medroxyprogesterone acetate, 5–10 mg daily for at least 10 days every month. Absence of induced bleeding requires workup. In the treatment of dysfunctional menometrorrhagia or polymenorrhea, progestins are prescribed for 10 days to 2 weeks (to induce stabilizing predecidual stromal changes) followed by a withdrawal flow — the so-called “medical curettage.” Thereafter, repeat progestin is offered cyclically for at least the first 10 days of each month to ensure therapeutic effect. Failure of progestin to correct irregular bleeding requires diagnostic reevaluation. If contraception is desired, the use of an oral contraceptive is a better choice.

Oral Contraceptive Therapy

In young women, anovulatory bleeding may be associated with prolonged endometrial buildup, delayed diagnosis, and heavy blood loss. In these cases, combined progestin-estrogen therapy is used in the form of combined oral contraceptives. Any of the low-dose oral combination monophasic tablets are useful. Whatever formulation is available or chosen, therapy is administered as one pill twice a day for 5–7 days. This therapy is maintained despite cessation of flow within 12–24 hours. If flow does not abate, other diagnostic possibilities (polyps, incomplete abortion, and neoplasia) should be reevaluated.

If flow does diminish rapidly, the remainder of the week of treatment can be given over to the evaluation of causes of anovulation, investigation of hemorrhagic tendencies, and blood replacement or initiation of iron therapy. In addition, the week provides time to prepare the patient for the estrogen-progestin withdrawal flow that will soon be induced. For the moment, therapy has produced the structural rigidity intrinsic to the compact pseudodecidual reaction. Continued random breakdown of formerly fragile tissue is avoided and blood loss stopped. However, a large amount of tissue remains to react to estrogen-progestin withdrawal. The patient must be warned to anticipate a heavy and severely cramping flow 2–4 days after stopping therapy. If not prepared in this way, it is certain that the patient will view the problem as recurrent disease or failure of hormonal therapy.

In successful therapy, on the 5th day of flow or in the usual Sunday start fashion, a low dose combination oral contraceptive medication (one pill a day) is started. This will be repeated for several (usually three) 3-week treatments, punctuated by 1-week withdrawal flow intervals. A decrease in volume and pain with each successive cycle is reassuring. Oral contraceptives reduce menstrual flow by at least 60% in normal uteri. Early application of the estrogen-progestin combination limits growth and allows orderly regression of excessive endometrial height to normal controllable levels. If the estrogen-progestin combination is not applied, abnormal endometrial height and persistent excessive flow will recur.

In the patient not requiring contraception, in whom cyclic estrogen-progestin for 3 months has reduced endometrial tissue to normal height, the oral contraceptive can be discontinued and unopposed endogenous estrogen permitted to reactivate the endometrium. In the absence of spontaneous menses, the recurrence of the anovulatory state is suspected, and a brief premenstrual course of an orally active progestin is administered to counter endometrial proliferation. Once pregnancy is ruled out, medroxyprogesterone acetate, 5–10 mg orally daily for at least 10 days, is given monthly. Reasonable flow (progestin withdrawal flow) will occur 2–7 days after the last pill. With this therapy, excessive endometrial buildup is avoided, and an increased risk of endometrial and possibly breast cancer is avoided. If contraception is desired, routine use of oral contraception is warranted and will also be of prophylactic value.

Depot-medroxyprogesterone acetate in the dose used for contraception, 150 mg intramuscularly every 3 months, is a useful option for poorly compliant patients. Breakthrough bleeding is treated with estrogen as discussed below.

Estrogen Therapy

Intermittent vaginal spotting is frequently associated with minimal (low) estrogen stimulation (estrogen breakthrough bleeding). In this circumstance, where minimal endometrium exists, the beneficial effect of progestin treatment is not achieved, because there is insufficient tissue on which the progestin can exert action. A similar circumstance also exists in the younger anovulatory patient in whom prolonged hemorrhagic desquamation leaves little residual tissue.
In these circumstances, when bleeding is acute and heavy, high-dose estrogen therapy is applied by using as much as 25 mg conjugated estrogen intravenously every 4 hours until bleeding abates or for 24 hours. This is the sign that the “healing” events are initiated to a sufficient degree. The mechanism of action for estrogen is believed to be a stimulus to clotting at the capillary level. Progestin treatment (usually an oral contraceptive) is started at the same time. Where bleeding is less, lower oral doses of estrogen (1.25 mg of conjugated estrogens or 2.0 mg estradiol daily for 7–10 days) can be prescribed initially. When bleeding is moderately heavy, a more intensive oral program can be utilized, 1.25 mg conjugated estrogens or 2 mg estradiol every 4 hours for 24 hours, followed by the single daily dose for 7–10 days. All estrogen therapy must be followed by progestin coverage and a withdrawal bleed.

Estrogen therapy is also useful in two examples of problems associated with progestin breakthrough bleeding. These are the breakthrough bleeding episodes occurring with use of oral contraception or with depot forms of postnatal agents. In the absence of sufficient endogenous and exogenous estrogen, the endometrium shrinks by pharmacologically induced pseudoatrophy. Furthermore, it is composed almost exclusively of pseudodecidual stroma and blood vessels with minimal glands. Peculiarly, experience indicates that this type of endometrium also leads to the fragility bleeding more typical of pure estrogen stimulation.

The usual clinical story is a patient on long-standing oral contraception who, after experiencing marked diminution or absence of withdrawal flow in the pill-free interval, begins to see breakthrough bleeding while on medication. Conjugated estrogens, 1.25 mg, or estradiol, 2.0 mg daily for 7 days during, and, in addition to, the usual birth control pill administration are effective. This treatment rejuvenates the endometrium and intermenstrual flow stops. Another frequently encountered problem is the progestin breakthrough bleeding experienced with chronic depot administration of progestin (Depo-Provera). This therapy is used not only for contraception, but also in the treatment of endometriosis and the prevention of menses during chemotherapy. In 75% of recipients, continuous therapy is not associated with abnormal menstrual bleeding. In the remainder, breakthrough progestin bleeding occurs. Judicious use of estrogen is the appropriate and effective therapy in these instances.

Bleeding problems are common with Norplant, another example of the effect of persistent postgestational influence on the endometrium. Patients who can no longer tolerate prolonged bleeding will benefit from a short course of oral estrogen as above.

The Risk Associated with Estrogen Therapy

There is concern that high doses of estrogen could precipitate a thrombotic event. More than one oral contraceptive per day and multiple doses of oral or intravenous estrogen in a 24-hour period certainly should be regarded as high doses. There are no data available, however, to verify or quantify any risk associated with this use of hormonal therapy. This treatment must be chosen by clinician and patient after weighing the risk-benefit considerations that surround the uterine bleeding problem. As a matter of clinical judgment and prudent practice, lower doses can be used in patients with lifestyle or medical history consistent with an increased risk of vascular complications. In women with a past episode or a positive family history of idiopathic venous thromboembolism, exposure to high doses of estrogen should be avoided.

The Use of Antiprostaglandins

There seems little doubt that prostaglandins (PG) have important actions on the endometrial vasculature and, presumably, on endometrial hemostasis. The concentrations of PGE_2 and PGF_2α increase progressively in human endometrium during the menstrual cycle, and nonsteroidal eicosanoid synthesis inhibitors decrease menstrual blood loss perhaps by also altering the balance between the platelet proaggregating vasoconstrictor thromboxane A_2 (TXA_2) and the antiaggregating vasodilator prostacyclin (PGI_2). Excessive bleeding in women with menorrhagia can be reduced by approximately 40–50%. In a study of ovulating women with menorrhagia, treatment during menses with a progestin-containing IUD, the progestin in combination with 0.35 mg norethindrone, was effective. In 30 patients treated for 6 months, the bleeding was reduced on average by 75% of control periods and in 93% of cycles in which the progestin was administered.

Treatment With Desmopressin

Therapy with a GnRH agonist can achieve short-term relief from a bleeding problem, for example, in a patient with renal failure or a blood dyscrasia. This choice is a good one for patients who experience menstrual bleeding problems after organ transplantation (especially after liver transplantation) where the toxicity of immunosuppressive drugs makes the use of sex steroids less desirable. However, the expense and long-term side effects make this an unlikely choice for chronic therapy. If long-term GnRH agonist therapy is chosen, after gonadal suppression is achieved (2–4 weeks), we recommend add-back treatment with a daily combination of 0.625 mg conjugated estrogens or 1.0 mg estradiol and 2.5 mg medroxyprogesterone acetate or 0.35 mg norethindrone.

Ablation of the Endometrium

Persistent bleeding despite treatment is both aggravating and concerning. Hysterec-tomy is an appropriate choice for some of these patients. Others would prefer to avoid a major operation, and still others have conditions that make major surgery a high-risk procedure. Patients and clinicians should consider the option of hysteroscopic endometrial ablation. Ablation of the endometrium can be accomplished with either a laser, a resectoscope with a loop or rolling ball electrode, or radiofrequency ablation. A new technique, ablation with a saline infusion, utilizes a saline ball to expose the endometrium to an 85°C temperature for 10–15 minutes using the circulation of hot saline. In a comparison with the rolling ball technique, the thermal balloon was as effective and with fewer side effects. In a randomized comparison of medical treatment of menorrhagia compared with hysteroscopic endometrial ablation, overall results were better with surgical management. Nevertheless, for most women, medical management is a better first approach (certainly in terms of cost).

Success with these methods is not 100%. Approximately 90% of women with menorrhagia will have an improvement following an ablation procedure; only 40–50% will become amenorrheic. Some women will continue to have menorrhagia. In a randomized trial comparing endometrial resection and hysterec-tomy, subsequent surgery was required by 22% of the women after endometrial resection compared with 9% in the hysterec-tomy group. The best results are obtained if the endometrium is first suppressed for 4–6 weeks with either a high dose of a progestin, GnRH agonist treatment, or danazol. Caution must be exercised regarding the possibility of excessive absorption of irrigating fluid with subsequent fluid overload. Despite the advantages of lower risk, fewer complications, and more rapid
recovery with endometrial resection, patients treated by hysterectomy tend to be more satisfied with the outcome.  

There is concern that obliteration of segments of the uterine cavity can allow isolated, residual endometrium to progress to carcinoma without recognition.  

Endometrial ablation is not recommended for patients who are considered to be at high risk for endometrial cancer.

Less Appealing Choices

Antifibrinolytic agents (e.g., tranexamine acid) are associated with a large number of side effects. Danazol effectively reduces menstrual blood loss, but treatment requires significant doses maintained for a long time period. The expense and androgenic side effects associated with danazol make it a poor choice.

Hyperplasia Versus Neoplasia

Ferenczy argues that there are two separate and biologically unrelated diseases: endometrial hyperplasia and endometrial neoplasia. He suggests that all hyperplasia without atypia be referred to as endometrial hyperplasia, and this is usually not a precursor of carcinoma. He further proposes that lesions with cytologic atypia be referred to as endometrial intraepithelial neoplasia (EIN). In these cases, persistence after multiple curettages or high-dose progestin therapy is approximately 75%. EIN would replace the following terms: atypical adenomatous hyperplasia and carcinoma in situ of the endometrium. This lesion is characterized by nuclear atypia of the cells lining the endometrial glands (enlargement, rounding, and pleomorphism of the nuclei with aneuploid DNA content). Invasive carcinoma is distinguished from EIN by stromal invasion.

EIN is best treated surgically: If future pregnancy is desired, daily progestin therapy (30 mg medroxyprogesterone acetate daily) should be followed by repeat endometrial aspiration curettage in 3–4 months. If EIN is still present, the choice is between surgery and high dose progestin (200 mg medroxyprogesterone acetate daily. 500 mg megestrol acetate biweekly, or depot medroxyprogesterone acetate 1000 mg weekly) with repeat biopsy surveillance.

The benign lesions include all of the following traditional interpretations: anovulatory, proliferative, cystic glandular hyperplasia, simple hyperplasia, adenomatous hyperplasia without atypia. These lesions are basically the same (perhaps exaggerations at most) as preovulatory, proliferative endometrium. This hyperplasia regresses spontaneously, after curettage, or with hormonal treatment. This argument is an important one: benign lesions can be treated hormonally. With one exception, only the presence of cytonuclear atypia should raise immediate and serious concern for progression to cancer. The exception is the patient with an apparently benign lesion that does not respond to progestin therapy. Was the diagnosis accurate? Will clinical data eventually indicate that this type of patient is also at higher risk for progression?

Summary of Key Points in Therapy of Anovulatory (Dysfunctional) Bleeding

1. Intense estrogen-progestin therapy for 7 days.
2. Cyclic low-dose oral contraceptive for 3 months.
3. If contraception is desired, continue oral contraception.
4. If not exposed to pregnancy, medroxyprogesterone acetate, 5–10 mg daily for at least 10 days every month.

If bleeding has been prolonged, if biopsy yields minimal tissue, if the patient is on progestin medication, or if follow-up is uncertain:

Conjugated estrogens (1.25 mg) or estradiol (2.0 mg) daily for 7–10 days, followed by the daily estrogen combined with 10 mg medroxyprogesterone acetate for 7 days. If acute bleeding is moderately heavy, the oral estrogen dose can be administered every 4 hours during the first 24 hours. For very heavy, acute bleeding, conjugated estrogen, 25 mg intravenously every 4 hours until bleeding stops or significantly slows, and then proceed to Step 1 above. If no response in 24 hours, proceed to D and C.

The clinical problem of dysfunctional bleeding is associated with either anovulation and estrogen withdrawal or breakthrough bleeding, or with anovulation caused by exogenous progestin medication and bleeding due to postgestational endometrial breakthrough. These categories of bleeding lack the three important characteristics of normal estrogen-progesterone withdrawal bleeding:

1. Universal, simultaneous change in all segments of the endometrium.
2. An orderly progression of events involving a rigid, compact structure.
3. Vasomotor rhythmicity with vasoconstriction, structural collapse, and clotting.

Questioning should be directed by the differential diagnosis of abnormal uterine bleeding. Clues to the diagnosis may be apparent on physical examination, such as hirsutism, acne, galactorrhea, thyroid enlargement, evidence of an eating disorder, bruises, and, of course, abnormalities on examination of the pelvis. Brown, dark colored bleeding is often secondary to obstruction in a Müllerian anomaly. Laboratory studies that can be helpful (but not always necessary) are coagulation studies (prothrombin time, partial thromboplastin time, platelet count, bleeding time, and the ristocetin cofactor assay for von Willebrand’s factor), quantitative human chorionic gonadotropin (HCG), prolactin, thyroid function tests, liver function tests, and appropriate cervical cultures. However, it is appropriate to move quickly to empirical hormonal therapy because pathology will eventually declare itself.

Office aspiration biopsy of the endometrium should always be performed in patients considered to be at high risk for endometrial hyperplasia and cancer. Texts and review articles continue to emphasize that endometrial biopsy is in order if the patient is older; e.g., greater than 35 or 40 years old. It is not the age of the patient that is critical; it is the duration of exposure to unopposed estrogen. Women in their 20s and even teenagers can develop endometrial cancer. The small flexible suction cannulas are preferred for greater patient comfort, and results are comparable with the older, traditional methods. Office hysteroscopy is also useful for the direction of biopsies and the detection of polyps and submucous myomas.

Therapy involves an initial choice between intensive estrogen-progesterin combination medication or relatively high doses of estrogen. The estrogen-progesterin combination will be ineffective unless endometrium of sufficient quantity and responsiveness to allow the formation of pseudodecidual tissue is present. Therefore, the initial choice of therapy should be estrogen in the following situations:

1. When bleeding has been heavy for many days and it is likely that the uterine cavity is now lined only by a raw basalis layer.
2. When the endometrial curet yields minimal tissue.
3. When the patient has been on progestin medication (oral contraceptives, or intramuscular progestins) and the endometrium is shallow and atrophic.
4. When follow-up is uncertain, because estrogen therapy will temporarily stop all categories of dysfunctional bleeding.

If estrogen therapy does not significantly abate flow within 24 hours, reevaluation is mandatory, and the need for curettage is likely. It is believed that patients with coagulation disorders respond better if the uterine cavity is first evacuated with a suction curet. Consideration should be given to the empirical administration of fresh
Once the acute bleeding episode in an anovulatory patient is under control, the patient should not be forgotten. With persistent anovulation, recurrent hemorrhage is a common sequel. Indeed, more importantly, chronic unopposed estrogen stimulation to the endometrium can eventually lead to atypical tissue changes. It is absolutely necessary that the patient undergo periodic pregestational withdrawal either with a routine oral contraceptive regimen or if contraception is not desired, a progesterational agent (medroxyprogesterone acetate, 5–10 mg daily for at least 10 days) should be administered every month.

Curettage is not the first line of defense, but, rather, the last. The use of appropriate steroids for the clinical management of dysfunctional bleeding is based on a physiologic understanding of the endometrium and its responses to hormones. Adherence to this program will avoid D and C except in rare cases of dysfunctional bleeding and except in those cases in which bleeding is due to a pathologic entity within the reproductive tract where A and C is truly indicated and necessary.

Remember to seek pathology if dysfunctional uterine bleeding is persistent or recurrent after hormonal treatment. 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 or careful fluoroscopic examination or hysteroscopy (especially with filling of the uterine cavity with fluid) may reveal a myoma or polyp; hysteroscopy can also direct a more accurate biopsy of the endometrium. A pathologic problem such as this should especially be suspected in the puzzling case of the patient who has abnormal bleeding and ovulatory cycles.

Patients who are ovulating but have a heavy menstrual flow (menorrhagia) can be effectively treated with prostaglandin inhibitors, progestins administered daily for 5 to 7 days, or oral contraceptives in the routine manner. If contraception is not required, we prefer the use of one of the fenamate prostaglandin inhibitors (which block both synthesis and prostaglandin receptors). Women who have menorrhagia but are ovulating should be evaluated for a coagulation disorder (prothrombin time, partial thromboplastin time, platelet count, bleeding time, and the ristocetin cofactor assay for von Willebrand's factor). The IUD that releases progesterone or a progestin is a good choice for women with menorrhagia, and it should be considered in patients with chronic illnesses.

Chapter References

11. Kirkland JL, Murthy L, Stancel GM. The frequency of inherited bleeding disorders in women with menorrhagia.
13. Falbo P, Marjamaa K, Lasansky A, Sainio K. Menorrhagia in women using oral contraceptives: fact or fiction?
26. Schatz F, Alperin S, Papp C, Toth-Pal E, Hauskenecht V, Lockwood CJ. Plasminogen activator activity during decidualization of human endometrial stromal cells is regulated by protrombin time, partial thromboplastin time, platelet count, bleeding time, and the ristocetin cofactor assay for von Willebrand's factor). The IUD that releases progesterone or a progestin is a good choice for women with menorrhagia, and it should be considered in patients with chronic illnesses.


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.

In this chapter, the factors involved in normal growth and development of the breast will be reviewed, including the physiology of normal lactation. A description of the numerous factors leading to inappropriate lactation will follow, and, finally, the endocrine aspects of breast cancer will be considered.

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 nonmuscular 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-a and ER-b. 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 influences growth of the alveolar components of the lobule. 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, prolacin, and growth hormone. Of course, the ubiquitous growth factors are also involved, but the molecular mechanisms remain to be determined. Nevertheless, experimental evidence in mice indicates that progesterone is the key hormone required for mammary growth and differentiation; estrogen is necessary because the synthesis of progesterone receptors requires the critical presence of estrogen.

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. 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. 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. Studies using tissue from reduction mammaplasties or from breast tissue near a benign or malignant lesion have demonstrated a peak in mitotic activity during the luteal phase. Using fine needle biopsy tissue, an immunocytochemical marker of proliferation was higher in the luteal phase than in the proliferative phase. 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. Therefore, breast and endometrium epithelial cells may be more similar than initially 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, 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 will begin 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. With one exception, hormone therapy is totally ineffective in producing a permanent change in breast shape or size. Of course in patients with primary amenorrhea due to deficient ovarian function, estrogen treatment will induce 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.

Accessory nipples can be found anywhere from the groin to the neck, remnants of the mammary line that extends early in embryonic life (6th week) along the ventral, lateral body wall. They occur in approximately 1% of women and require no therapy.

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 the short arm of chromosome 6, producing a molecule that in its major form is maintained in 3 loops by disulfide bonds. Simultaneous measurements of prolactin by both bioassay and immunoassay reveal discrepancies. 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, post translational modification, and peripheral metabolism. 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 moieties can produce differences in biologic activity and immunoreactivity. However, the more glycosylated form of prolactin is the predominant form of prolactin secreted into the circulation. Modification of prolactin also includes phosphorylation, deamidation, and sulfation. Little prolactin probably results 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.

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 development. Pit-1 also 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. 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 also influences prolactin production by suppressing dopamine secretion.
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. 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. Suppression involves the stimulation of endogenous dopamine. Suppression of prolactin receptor expression.

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. 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. There is marked variability in maternal prolactin levels in pregnancy, with pulsatile secretion and a durnal variation similar to that found in nonpregnant subjects. The peak level occurs 4–5 hours after the onset of sleep.

Made by the placenta and actively secreted into the maternal circulation from the 6th week of pregnancy, human placental lactogen (HPL) rises progressively to reach a level of approximately 6000 ng/mL at term. HPL, though displaying less activity than prolactin, is produced in such large amounts that it may exert a lactogenic effect.

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 prolactin parallels 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 role for the placenta in prolactin synthesis. The placenta contains prolactin receptors, and prolactin is believed to be a placental hormone.

Prolactin receptors are expressed in many tissues throughout the body. Various decidual factors regulate prolactin synthesis and release, including relaxin, insulin, and insulin-like growth factor-I (discussed in Chapter 4). Prolactin produced in extraplacental 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 biological properties of decidual prolactin are identical to those of pituitary prolactin. It is hypothesized that amniotic fluid prolactin produces 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 steehead 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.

**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. 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. There is marked variability in maternal prolactin levels in pregnancy, with pulsatile secretion and a durnal variation similar to that found in nonpregnant subjects. The peak level occurs 4–5 hours after the onset of sleep.

Made by the placenta and actively secreted into the maternal circulation from the 6th week of pregnancy, human placental lactogen (HPL) rises progressively reaching a level of approximately 6000 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. In the mouse, inhibition of milk protein production is due to progesterone suppression of prolactin receptor expression. The effective use of high doses of estrogen to suppress postpartum lactation indicates that pharmacologic amounts of estrogen also block prolactin action.
The principal hormone involved in milk biosynthesis is prolactin. Without prolactin, synthesis of 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 non-pregnant levels in a non-breastfeeding 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 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.

Reports of hypertension, seizure, myocardial infarctions, and strokes associated with its postpartum use.

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 two-fold increase that is essential for continuing milk production. The pattern or values of prolactin levels does not predict the postpartum duration of amenorrhea or infertility. 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). Prolactin sustains the secretion of casein, fatty acids, lactose, and the volume of secretion, while oxytocin contracts myoepithelial cells and emplies 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. 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. However, bone density rapidly returns to baseline levels in the 6 months after weaning. 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 1500 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. 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.

Antibodies are present in breast milk and contribute to the health of an infant. 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 virus are advised not to breastfeed. Vitamin A, vitamin B12, 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 non-insulin-dependent diabetes mellitus, partly because overfeeding and excess weight gain are more common with bottlefeeding.

Frequent emptying of the lumen is important for maintaining an adequate level of secretion. Indeed, after the 4th 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.

The ejection of milk from the breast does not occur 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 openings in the nipple. This rapid release of milk is called “letdown.” This important role for oxytocin is evident in knockout mice lacking oxytocin who undergo normal parturition, but fail to nurse their offspring.

In many instances, the activation of oxytocin release leading to letdown 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 sucking. 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. Successful breastfeeding can be achieved by approximately half of the women by ingestion of 25 mg chlorpromazine tid together with vigorous nipple stimulation every 1–3 hours. Milk production will not appear for several weeks. This preparation ideally should be practiced for several months.

Cessation of Lactation

Lactation can be terminated by discontinuing suckling. The primary effect of this cessation is loss of milk letdown 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 letdown 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, seizure, 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 level of nutrition of the mother (if low, the longer the contraceptive interval), the intensity of suckling, and the extent to which supplemental food is added to the infant diet. 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 (96% efficacy); with menstruation or after 6 months, the chance of ovulation is approximately 25%, and with full or nearly full breastfeeding, approximately 70% of women remain amenorrheic through 6 months and only 37% through one year; nevertheless with exclusive breastfeeding, the contraceptive efficacy at one year is high, at 92%. Fully breastfeeding women commonly have some vaginal bleeding or spotting in the first 8 postpartum weeks, but this bleeding is not due to ovulation.

Supplemental feeding increases the chance of ovulation (and pregnancy) even in amenorrheic women. Total protection is achieved by the exclusively breastfeeding woman for a duration of only 10 weeks. Half of women studied who are not fully breastfeeding ovulate before the 6th week, the time of the traditional postpartum visit; a visit during the 3rd 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 nonbreastfeeding women, gonadotropin levels remain low during the early puerperium and return to normal concentrations during the 3rd to 5th 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 ovulation occurred before 25 days after delivery. Of the 22 women, however, 11 ovulated before the 6th postpartum week, underscoring the need to move the traditional postpartum medical visit to the 3rd 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 2nd week postpartum.

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. Prolactin concentrations are increased in response to the repeated sucking stimulus of breastfeeding. Given sufficient intensity and frequency, prolactin levels will remain elevated. Under these conditions, follicle-stimulating hormone (FSH) concentrations are in the normal range (having risen from extremely low concentrations at delivery to follicular range in the 3 weeks postpartum) and luteinizing hormone (LH) values are in the low normal range. Despite the presence of gonadotropin, the ovary, during lactational hyperprolactinemia, does not display follicular development and does not secrete estrogen.

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 trophic hormone stimulation.

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. 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. 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. 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 amenorrhea, lactating women with prolactin GnRH fully restores pituitary secretion and normal ovarian cyclic activity.

At weaning, as prolactin concentrations fall to normal, gonadotropin concentrations increase, and estradiol secretion rises. This prompt resumption of ovarian function is also indicated 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 non-physiologic in that it is inappropriate (not immediately related to pregnancy or the needs of a child), persistent, and sometimes excessive. Although usually white or yellow, the color may be yellow or even green. In the latter circumstance, breast disease (such as cancer) 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. 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 one breast. Amenorrhea does not necessarily accompany galactorrhea, even in the most serious provocative disorders.

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 may be a consequence of prolactin elaboration and secretion from pituitary tumors which function independently of the otherwise appropriate restraint 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 PIF. 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, a-methyltyrosine, 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 in 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 phenothiazines, but essentially never as high as 100 ng/mL. Approximately 30–50% will exhibit galactorrhea that should not persist beyond 3–6 months after drug treatment is discontinued.

3. Exogenous estrogens or androgens can be converted to estrogen by the liver with galactorrhea occurring when the estrogen level exceeds the 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 PIF and release of pituitary prolactin, and direct stimulation of the pituitary lactotropes. Galactorrhea developing during oral contraceptive administration may be most noticeable during the days free of medication (when the steroids are cleared from the body and the prolactin interfering action of the estrogen and progesterin on the breast wages). Galactorrhea caused by excessive estrogen disappears within 3–6 months after discontinuing medication. This is now a rare occurrence with the lower dose pills. A longitudinal study of 120
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.

5. Prolonged intensive suckling can also release prolactin, via hypothalamic reduction of PIF. Similarly, thoracotomy scars, cervical spinal lesions, and herpes zoster can induce prolactin release by activating the afferent sensory neural arc, thereby simulating suckling.

6. Stresses can inhibit hypothalamic PIF, 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 PIF to the pituitary) allow release of excess prolactin leading to galactorrhea.

8. Increased prolactin concentrations can result from non-pituitary sources such as lung 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 have been applied to variants of the lactation syndromes. These were based on the association of galactorrhea with intrasellar tumor (Forbes, Henneman, Griswold, and Albright, 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 is neither 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: oligoovulation, corpus luteum insufficiency, as well as amenorrhea. About one-third of women with secondary amenorrhea will 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 DHA (dehydroepiandrosterone) and its sulfate (DHAS) or is primarily related to the chronic anovulation of these patients (and hence ovarian androgen secretion) is not settled.

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 immunoactivity (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 workup of hyperprolactinemia is presented in detail in Chapter 11, "Amenorrhea." It is worth reemphasizing the salient clinical issues here.

With the current diagnostic techniques there is no difficulty in discovering and monitoring the size and function of a pituitary prolactin-secreting tumor. With few exceptions the combination of elevation in basal levels of prolactin and radiographic imaging offers complete confidence in diagnosing sellar pathology. The major concern remains in determining management—medical, surgical, or expectant? The considerations that influence management include:

1. Microadenomas, if exclusively prolactin-producing, rarely progress to macroadenoma size. Most are exceedingly slow-growing or stable.
2. Some tumors regress spontaneously. Medical therapy with dopamine agonists shrinks tumors and can prevent growth, although complete elimination of a tumor
by dopamine agonist treatment does not occur, and rapid regrowth usually follows discontinuation of the drug.

3. Transsphenoidal microsurgery is a very safe procedure, but there is a high recurrence rate.

4. The presence of a prolactin-secreting adenoma does not represent a contraindication to pregnancy or the use of exogenous hormones such as oral contraceptives.

As a result of these considerations, many patients can be observed, others treated medically, and, rarely, some treated with surgery, with or without prior medically-induced tumor reduction (see Chapter 11).

Treatment of Galactorrhea

Galactorrhea as an isolated symptom of hypothalamic dysfunction existing in an otherwise healthy woman does not require treatment. Periodic prolactin levels will, 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, therefore, is the therapy of choice. Even with normal prolactin concentrations and normal imaging, treatment with a dopamine agonist can eliminate galactorrhea.

We have adopted a conservative approach of close surveillance for pituitary prolactin-secreting adenomas, recommending surgery only for those tumors that display rapid growth or those tumors that are already large and do not shrink in response to dopamine agonists. If the prolactin level is greater than 100 ng/mL, or if the coned-down view of the sella turcica is abnormal, we recommend magnetic resonance imaging (MRI) evaluation. If the MRI rules out an empty sella syndrome or a suprasellar problem, surgical intervention after preoperative dopamine agonist treatment is then dictated by the patient's desires, the size of the tumor, and the response of the tumor to a dopamine agonist. In patients with prolactin levels less than 100 ng/mL and with normal coned-down views of the sella turcica, an annual prolactin level and periodic coned-down views are indicated for continued observation to detect a growing tumor. Dopamine agonist therapy is recommended for patients wishing to achieve pregnancy, and for those patients who have galactorrhea to the point of discomfort.

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. Approximately 70% of women report premenstrual breast discomfort in surveys, and interference with activities is recorded in 10–30%.

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. A good practice is to start with 5 mg methyltestosterone every other day during the time of discomfort. In recent years, however, 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. 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.

Clinical observations have 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 have failed to demonstrate a link between methylxanthine use and mastalgia, mammographic changes, or atypia (premalignant tissue changes).

Cancer of the Breast

Scope of the Problem

Currently, female American newborns have a lifetime probability of developing breast cancer of 12.5%, about 1 in 8, double the risk in 1940. The incidence increased over the past 4 decades but plateaued in 1987 (about 180,000 new cases of invasive breast cancer per year). Mortality rates remained disappointingly constant (44,000 deaths per year) until a decline began in 1990. The 5-year survival rate for localized breast cancer has risen from 72% in the 1940s to 97%.

Clinical observations have 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 have failed to demonstrate a link between methylxanthine use and mastalgia, mammographic changes, or atypia (premalignant tissue changes).
The breast is the leading site of cancer in U.S. women (29.7% of all cancers) and is now, unfortunately, (because smoking is obviously the reason) exceeded by lung cancer as the leading cause of death from cancer in women.

Breast cancer has an increasing frequency with age. Nearly 77% of all breast cancers occur in women over age 50; only 6.5% of all cases occur under age 40, 15% under age 50. A woman at age 70 has almost 16 times the risk as a 40-year-old woman. However, after menopause, although the incidence rate continues to increase, it is less dramatic than the rise before menopause. Over the age of 50, American white women are more likely to develop breast cancer than black American women; under the age of 50, the opposite is true. The 5-year survival rate for women who develop breast cancer under age 45 is 79% compared to 87% for women ages 65 and older.

The Chances of Developing Breast Cancer According to Age

| Age | Probability  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>By age 25</td>
<td>1 in 19,608</td>
</tr>
<tr>
<td>By age 30</td>
<td>1 in 2,525</td>
</tr>
<tr>
<td>By age 35</td>
<td>1 in 622</td>
</tr>
<tr>
<td>By age 40</td>
<td>1 in 217</td>
</tr>
<tr>
<td>By age 45</td>
<td>1 in 93</td>
</tr>
<tr>
<td>By age 50</td>
<td>1 in 50</td>
</tr>
<tr>
<td>By age 55</td>
<td>1 in 33</td>
</tr>
<tr>
<td>By age 60</td>
<td>1 in 24</td>
</tr>
<tr>
<td>By age 65</td>
<td>1 in 17</td>
</tr>
<tr>
<td>By age 70</td>
<td>1 in 14</td>
</tr>
<tr>
<td>By age 75</td>
<td>1 in 11</td>
</tr>
<tr>
<td>By age 80</td>
<td>1 in 10</td>
</tr>
<tr>
<td>By age 85</td>
<td>1 in 9</td>
</tr>
<tr>
<td>Lifetime</td>
<td>1 in 8</td>
</tr>
</tbody>
</table>

Over the years, breast cancer has continued its 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. The survival rate is higher with axillary lymph nodes negative for disease compared with positive nodes. Because of this recognition for 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. 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.

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. In fact, women over the age of 30 years at the time of their first birth have a greater risk than women who never become pregnant. 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). There is, however, a significant protective effect with increasing parity, present even when adjusted for age at first birth and other risk factors. 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.

Data from the Nurses’ Health Study indicate that a positive family history for breast cancer overwhelms the protective effect of pregnancy, and women with a mother or sister history of breast cancer actually had an increase in risk enhanced by first pregnancy.
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. This increase may reflect 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. 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). It is argued that breast cells which have already begun malignant transformation are adversely affected by the hormones of pregnancy, while normal stem cells become more resistant.

Conflicting results have been reported in over 20 studies examining the risk of breast cancer associated with the number of abortions (especially induced abortions) experienced by individual patients. 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 has been 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. 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 which 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 estril, a decrease in dehydroepiandrosterone and dehydroepiandrosterone sulfate, and lower prolactin levels all persist for many years after delivery. 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% reduced risk) only for premenopausal breast cancer. There is a unique and helpful study of the Chinese Tanka, who are boat people living on the coast of southern China. 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. 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. The impact of lactation, if significant, must be small. In both cohort and case-control studies, there is good evidence that cosmetic breast augmentation does not increase the risk of breast cancer. Specifically, studies have failed to indicate an increased risk of breast cancer in women who have had cosmetic breast implants.

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 ovariecctomized. 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.

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 dilatation of terminal ducts (duct ectasia) and alveoli. There is good reason to eliminate the phrase "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. The most important variable on biopsies is the degree and character of the epithelial proliferation. Women with atypical hyperplasia had a relative risk of 3.5, 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. Only 4–10% of benign biopsies have atypical hyperplasia. The point is that we needlessly frighten patients with the use of the phrase 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.

### Classification of Breast Biopsy Tissue According to Risk for Breast Cancer

<table>
<thead>
<tr>
<th>No increased risk:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosis</td>
<td></td>
</tr>
<tr>
<td>Duct ectasia</td>
<td></td>
</tr>
<tr>
<td>Fibroadenoma</td>
<td></td>
</tr>
<tr>
<td>Fibrosis</td>
<td></td>
</tr>
<tr>
<td>Mild hyperplasia (3–4 cells deep)</td>
<td></td>
</tr>
<tr>
<td>Mastitis</td>
<td></td>
</tr>
<tr>
<td>Periductal mastitis</td>
<td></td>
</tr>
<tr>
<td>Squamous metaplasia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slightly increased risk (1.5–2.0 times):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate or florid hyperplasia</td>
</tr>
<tr>
<td>Papilloma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk increased 3–5 times:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atypical hyperplasia</td>
</tr>
</tbody>
</table>

### 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. In data from the Centers for Disease Control and Prevention (CDC), these relative risks were observed.

| Affected mother or sister: | 2.3 relative risk. |
| Affected aunt or grandmother: | 1.5 relative risk. |
| Affected mother and sister: | 14.0 relative risk |

Results from the Nurses' Health Study indicate that the magnitude of risk associations with a positive family history is smaller than previously believed, comparable to the CDC data above, except for a relative risk of only 2.3 when both mother and sister (first-degree relatives) had breast cancer. In general, the size of the increased risk with a positive family history is approximately 2 times the normal incidence. In the Nurses' Health Study, there was no interaction between family history and the risk of breast cancer.
and alcohol intake; however, a positive mother or sister history of breast cancer eliminated the protective effects of late menarche, multiple births, and early births.

Using the data from the Cancer and Steroid Hormone Study, the CDC has provided tables which can be used for the purpose of counseling women with a family history of breast cancer. These tables estimate the risks according to the various combinations of affected first- and second-degree relatives and the ages at onset of breast cancer for the affected relatives. The risk is higher with early age of onset and more than one relative with breast cancer. For example, the predicted lifetime risk for a woman with two first-degree relatives (sisters and mother) who developed breast cancer before age 40 is approximately 44%, and with one first-degree relative with breast cancer between ages 30 and 39 years, 17%.

The breast and ovarian cancer gene (BRCA1) associated with familial cancer is on the long arm of chromosome 17, localized to 17q12–q21. Although other genetic alterations have been observed in breast tumors, mutations in BRCA1 are believed to be responsible for approximately 45% of familial breast cancer and 80% of families with both early onset breast and ovarian cancer. Inheritance of this gene can be either maternal or paternal; male carriers are at increased risk for colon and prostate cancers. A second locus, 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. Together, BRCA1 and BRCA2 account for 80% of families with multiple cases of early-onset breast cancer. About 5–10% of women who develop ovarian cancer have mutations in BRCA1.

BRCA1 encodes an 1863 amino acid protein with a zinc finger domain that is probably a tumor suppressor important in DNA transcription. Mutations in many different regions of the BRCA1 gene cause a loss or reduction in its function. Because not every individual with a mutation in this gene develops cancer, other factors are probably present which are needed for cancer initiation and progression. It is generally believed that widespread screening for women with BRCA1 mutations is difficult, as the gene is usually not found in affected women and sometimes the BRCA1 gene can be so difficult to identify. Providing accurate numbers is a difficult task, because breast cancer has a multifactorial etiology with both genetic and environmental factors. The BRCA1 gene may also 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.

High-risk families have a high probability of harboring a mutation in a dominant breast cancer susceptibility gene, transmitted by either parent, and accounting for approximately 5% of breast cancer in the general population. 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). Among women of Ashkenazi Jewish descent, the prevalence of BRCA1 and BRCA2 mutations is about 2%. 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. Later more representative studies revealed a lower prevalence, as low as 3% in the general population. 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 3 or more cases of breast cancer within a family are strong predictors of BRCA1 mutations. Thus, genetic screening should be reserved for patients with high-risk families.

Family History Characteristics Associated With the Presence of BRCA1

Early age of onset of breast cancer within a family. Relatives with ovarian cancer. 3 or more relatives with breast cancer. Ashkenazi 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 3 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. 

<table>
<thead>
<tr>
<th>Affected Relative</th>
<th>Age of Affected Relative</th>
<th>Cumulative Breast Cancer Risk by Age 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>One first-degree relative</td>
<td>&lt; 50 years old</td>
<td>13–21%</td>
</tr>
<tr>
<td>50 or more years old</td>
<td>9–11%</td>
<td></td>
</tr>
<tr>
<td>One second-degree relative</td>
<td>&lt; 50 years old</td>
<td>10–14%</td>
</tr>
<tr>
<td>50 or more years old</td>
<td>8–9%</td>
<td></td>
</tr>
<tr>
<td>Two first-degrees relatives</td>
<td>Both &lt; 50 years old</td>
<td>35–45%</td>
</tr>
<tr>
<td>Both 50 years or older</td>
<td>11–24%</td>
<td></td>
</tr>
<tr>
<td>Two second-degrees relatives</td>
<td>Both &lt; 50 years old</td>
<td>21–26%</td>
</tr>
<tr>
<td>but both parental or maternal</td>
<td>Both 50 years or older</td>
<td>9–16%</td>
</tr>
</tbody>
</table>

BRCA1 and BRCA2 direct DNA screening is now possible. Each child of a carrier has a 50% chance of inheriting the mutation. Although it is recognized that this may be an over-estimation, the studies indicate that women who are carrying the BRCA1 mutation have an 87% cumulative lifetime risk of developing breast cancer—about 20% by age 40, 51% by age 50, and 85% by age 70, and a 20–65% risk for ovarian cancer (closer to 20% in more recent studies of less highly selected groups). In addition, the male relatives who are carrying this mutation have approximately a 3-fold increased risk of prostate cancer and 4-fold increased risk of colon cancer. The breast cancer risk is similar for women with BRCA2 mutations. 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, may have a better survival.

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.

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. When 3 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 all on the father’s side. The family presence of just one case of ovarian cancer further increases the likelihood of the BRCA1 mutation. Identifying the families that carry the BRCA2 gene uses the same historical criteria as that for the BRCA1 gene. In contrast to BRCA1 families, BRCA2 families have only a moderately increased incidence of ovarian cancer.

P=Identification 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.

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.

A problem exists when it comes to deciding what to do. While we have good data indicating prevalence and risk in these high-risk families, we do not have studies telling us what the risk reduction is from prophylactic surgery, nor do we know the efficacy of increasing surveillance programs for these identified individuals. 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. Treatment also applies with equal success to women who undergo prophylactic removal of the bulk of the tissue at risk should be effective. The first follow-up study of high risk women who have undergone prophylactic mastectomy indicated a major reduction in the number of breast cancers in the mastectomy group, although total prevention was not achieved.

Current recommendations from experts in this field are as follows. For an individual identified to be at high risk, clinical breast examination and mammography...
are demonstrated every 6–12 months beginning between ages 25 and 35. Evaluation every 6 months is appropriate because the BRCA1-related tumors have been demonstrated to be faster growing tumors. Although the evidence for the precise degree of protection offered by prophylactic mastectomy is not available, support should be provided for those women who choose this option. Pelvic examination, serum CA-125 levels, and transvaginal ultrasonography with color Doppler are recommended every 6–12 months for women under age 40. Prophylactic oophorectomy is recommended at the completion of child-bearing, preferably before age 35. It is of benefit to take oral contraceptives to reduce the ovarian cancer risk. 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). The magnitude of this protection is identical with that previously observed in the general population. In our view, postmenopausal estrogen therapy is not contraindicated for these women (see Chapter 18).

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. Lean women, however, have an increased incidence of breast cancer, although this increase is limited to small, localized, and well-differentiated tumors. Furthermore, studies have failed to find evidence for a positive relationship between breast cancer and dietary total or saturated fat or cholesterol intake. On the other hand, there is evidence that dietary fat is a stronger risk factor for postmenopausal breast cancer than for premenopausal breast cancer. Although a cohort study indicated that dietary fat is a determinant of postmenopausal breast cancer, the association did not achieve statistical significance. 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 intra-abdominal fat (android obesity) and the risk of breast cancer, a consequence of excessive caloric consumption, however, not a specific dietary component. Presumably, the connection between android obesity and breast cancer is through the metabolic perturbations 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. The common denominator may be the peripheral insulin resistance and hyperinsulinemia that become prevalent with aging in affluent, modern societies.

This specific metabolic change is becoming a common theme in various clinical conditions, particularly non-insulin-dependent diabetes mellitus, anovulation and polycystic ovaries, hyperpertension, and dyslipidemia. Hyperinsulinemia is found more often in women with breast cancer. In this subgroup of women, the risk of breast cancer may be unfavorably influenced by hormonal changes (such as a decrease in sex hormone-binding globulin and an increase in unbound, free estradiol). There are, indeed, many reasons to avoid excess body weight. The risk of breast cancer is reduced in women who exercise regularly.

In women, soy consumption produces a reduction in the circulating levels of estradiol. 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 another case-control study indicated a reduction in the risk of breast cancer, in women with a high consumption of soy and other legumes. It is by no means certain, however, that there is a direct effect of soy intake. Soy intake may be a marker for other factors in lifestyle or diet that are protective.

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. Evidence indicates that this reflects a change in diet and lifestyle, with an increase in risk associated with a gain in height and weight. 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, excess weight is associated with either an unchanged or slightly increased risk. 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 fewer estradiol levels (lower sex hormone-binding globulin), and a slightly greater risk for breast cancer. A large Swedish case-control study suggested that the principal factor is weight gain during adulthood, and that the impact on breast cancer emerges 10 years after menopause. 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.

Alcohol in the Diet

There is a modest increase in the risk for breast cancer with the consumption of one or more alcoholic drinks of all forms per day. Almost all of many studies conclude that two drinks daily increase the risk by about 30–40%. It is tempting to speculate 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 on alcohol intake. However, when alcohol ingestion is administered under experimental conditions, circulating estrogen concentrations are raised to high levels. 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.

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. 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. Measurement of this 17-ketosteroid 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.

Endogeneous Estrogen

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 provocatio, 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. 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 pregnancy. Women having had an early pregnancy continue to excrete more estradiol 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. 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 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 equal to that of estradiol, it can produce a similar biologic response. In pregnancy, where the concentration of estradiol is very great, it can be an important hormone, not just a marker. It is higher in breast cancer and there is a selective increase in breast cancer rather than the ratio of estradiol to estriol, a range rarely encountered either physiologically or pharmacologically. Below this range, estradiol is unimpeded, above this range estriol itself exerts estrogenic activity.

Stanley G. Korenman promulgated a most interesting hypothesis concerning the endocrinology of breast cancer. Recognizing that the endocrine changes thought to be related to the promotion or provocatio of breast cancer were small, inconsistent, and could hardly account for the differential risk of breast cancer
among populations, Korenman concluded that endocrine disorder is related to breast cancer by influencing the patient's susceptibility to environmental carcinogens.Called the “Open Window Hypothesis,” Korenman argued that unopposed estrogen stimulation is the most favorable state for tumor induction (the “open” window). Susceptibility to breast cancer declines with the establishment of normal luteal phase progesterone secretion and becomes very low during pregnancy; the open window is closed.

The two main open window periods are the pubertal years prior to the establishment of regular ovulatory menstrual cycles and the perimenopausal period of waning follicle maturation and ovulation. The prolongation of these open windows by obesity, infertility, delayed pregnancy, earlier menarche, and later menopause would be associated with greater susceptibility. This argument is supported by observational studies indicating that anovulatory and infertile women (exposed to less progesterone) have an increased risk of breast cancer later in life. However, the statistical power of these observational studies was limited by small numbers (all fewer than 15 cases).

Although theoretically appealing on the basis of presumed correlation with epidemiologic risks (infertility, late menopause) clinical research has not always confirmed the thesis. Young women at high genetic risk for breast cancer had normal luteal phases, and a group of premenopausal women with breast cancer also had normal luteal phases. Others have failed to find a link between anovulation and the risk of breast cancer. Indeed, women with the most irregular menstrual cycles (and presumably anovulatory) in the Nurses’ Health Study appeared to have a reduced risk of breast cancer.

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 with the incidence of breast cancer could be demonstrated. 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. 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. 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.

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). 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; however, no association could be demonstrated with percent free or bioavailable levels of estradiol, androstenedione, testosterone, or dehydroepiandrosterone. 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. 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.

The biologic plausibility and epidemiologic support for an estrogen link are impressive arguments. Whether the important factor is the total amount of estrogen, the proportion of estrogen metabolites, the ratio of estrogen to androgen, or the ratio of estrogen to sex hormone-binding globulin is not clear. However, epidemiologic studies indicate that with increasing duration of exposure, progesterone can limit breast epithelial growth as it does with endometrial epithelium. In vitro studies of normal breast epithelial cells reveal that progestins inhibit proliferation. Human breast tissue specimens removed after the patients were treated with estradiol and progestrone indicate that progestrone inhibits in vivo estradiol-induced proliferation. As noted, women who ultimately develop breast cancer do not have consistent blood levels of progesterone. 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 (discussed with references in Chapter 18). Medroxyprogesterone acetate is not associated with an increased risk of breast cancer when used for contraception over long durations.

Endogenous Progesterone

Because mitotic activity in the breast reaches its peak during the progesterone dominant luteal phase of the menstrual cycle, it can be 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. 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. In vitro studies of normal breast epithelial cells reveal that progestins inhibit proliferation. Human breast tissue specimens removed after the patients were treated with estradiol and progesterone indicate that progesterone inhibits in vivo estradiol-induced proliferation. As noted, women who ultimately develop breast cancer do not have consistent blood levels of progesterone. 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 (discussed with references in Chapter 18). Medroxyprogesterone acetate is not associated with an increased risk of breast cancer when used for contraception over long durations.

Exogenous Estrogen and Progestin

The relationship between the postmenopausal use of exogenous estrogens and the risk of breast cancer has been intensively studied (reviewed with complete references in Chapter 18). At the present time there is no conclusive evidence that estrogen doses known to protect against osteoporosis and cardiovascular disease (0.625 mg conjugated estrogens, 1.0 mg estradiol, or equivalent doses of other products) increase the risk of breast cancer. Some have concluded that a slight increase is noted with long duration of use; however, notable studies have failed to document such an increase. The extensive literature on this subject can be summarized as follows:

Some epidemiologic case-control and cohort studies conclude that long-term (5 or more years) of current use of postmenopausal hormone therapy is associated with a slight increase in the risk of breast cancer. This conclusion might be due to confounding biases, particularly detection and surveillance bias.

All epidemiologic studies fail to find an increased risk of breast cancer associated with short-term (less than 5 years) use or past use of postmenopausal hormone therapy.

The epidemiologic data agree that the addition of a progestin to the treatment regimen neither increases nor decreases the risk observed in individual studies.

The epidemiologic data indicate that a positive family history of breast cancer should not be a contraindication to the use of postmenopausal hormone therapy.

Women who develop breast cancer while using postmenopausal hormone therapy have a reduced risk of dying from breast cancer. This is because of two factors: (1) Increased surveillance and early detection; and (2) Acceleration of tumor growth so that tumors appear at a less virulent and aggressive stage.

Thyroid, Prolactin, Various Nonestrogen Drugs

Despite isolated suggestions of increased risk, hypothyroidism, reserpine, and prolactin excess, whether spontaneous or drug-induced, are not related to an enhanced risk of breast cancer.

Oral Contraception and Breast Cancer

The large number of women taking or having taken oral contraceptive steroids, combined with the belief that 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 have indicated no significant differences between 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.
Long-term use of oral contraception during the reproductive years is not associated with a significant increase in the risk of breast cancer that occurs later in life, after age 45. There is the possibility that a subgroup of young women who use oral contraceptives early and for more than 4 years has a slightly increased risk (a relative risk of less than 1.2) of breast cancer that occurs earlier in life, before age 45. The re-analysis of data from 54 studies indicated that breast cancers that occurred after ages 50 in women who were previous users of oral contraceptives were more localized, with less metastatic disease. Indeed, the risk of metastatic disease in ever users was significantly reduced. Breast cancer before age 45 is relatively rare, and furthermore, the results of these studies may reflect the same problem observed in studies with postmenopausal hormone therapy, detection/surveillance bias. The reduced risk of later metastatic disease and the increased appearance early in life suggest a situation analogous to pregnancy. Thus, breast cells that have already begun malignant transformation may be accelerated in their growth by exposure to oral contraceptives and appear early in localized form. With the accumulation of greater numbers of older women previously exposed to oral contraceptives, a protective effect may become evident. Overall, the message is reassuring. The statistical effect, if real, is small, and the data are also consistent with explanations other than an etiologic stimulation.

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 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. 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 uncertain whether this same protection is provided by the current low-dose formulations. A French case-control study has indicated a reduction of non-proliferative benign breast disease associated with low-dose oral contraception used before a first full-term pregnancy, but no effect on proliferative disease with use after a pregnancy.

Breast Cancer in Diethylstilbestrol (DES)-Exposed Women

From 1940 to 1970, diethylstilbestrol (DES), a potent synthetic estrogen, was prescribed in high doses for a variety 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. A large collaborative study, involving approximately 6000 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. A longer follow-up (more than 30 years) of this large cohort of DES-exposed women is now available. Exposure to DES is associated with a significant, but modest (less than two-fold), increase in the risk of breast cancer. 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. Certainly it is wise to recommend to DES-exposed women that they adhere religiously to screening for breast cancer, including mammography as discussed below.

Receptors and Clinical Prognosis

There is an excellent correlation between the presence of estrogen receptors and certain clinical characteristics of breast cancer, including response to endocrine therapy. Premenopausal and 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.

It appears that patients with estrogen receptors are those with the more slowly growing tumors. Several reports indicate that 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. The best prognosis is seen in patients with positive progesterone receptors, even with subclinical disease if the recurrent disease is still progesterone receptor positive. The loss of progesterone receptors is an ominous sign.

Tamoxifen and Breast Cancer

The purpose of adjuvant therapy of breast cancer therapy 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. Tamoxifen, in binding to the estrogen receptor, competitively inhibits estrogen binding. In vitro, the estrogen binding affinity for its receptor is 100–1000 times greater than that of tamoxifen. Thus, tamoxifen must be present in a concentration 100–1000 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. In addition, when bound to the estrogen receptor, tamoxifen prevents gene transcription by the TAF-2 pathway. In vitro studies demonstrate that these actions are 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. Adjuvant treatment with the antiestrogen, tamoxifen, achieves highly significant reductions in recurrence and increases in survival. The beneficial effect of tamoxifen is 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 occurs in the first 5 years, but continued impact on survival occurs throughout 10 years. A greater effect is seen with longer treatment (5 years). The 10-year survival difference is even greater that at 5 years. Adjuvant treatment (which is either tamoxifen, chemotherapy, or ovarian ablation) yields worldwide an extra 100,000 10-year survivors. There is an increased survival at 5 years of approximately 20%, 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. However, the results have indicated that there is little reason to extend tamoxifen treatment of breast cancer patients beyond 5 years. 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.

The major disturbing side effect is an increase in hot flushing. In a preliminary report from the prevention trial in England and in the U.S. trial, tamoxifen treatment of postmenopausal women prevented bone loss, but premenopausal women treated with tamoxifen had significant reductions in bone mineral density. Blurred and decreased vision has been reported associated with retinal changes, but it is uncertain whether this is a cause and effect relationship. In a prospective study of 63 patients in Greece, 6.3% developed retinopathy which was reversible except for retinal opacities. In the 2673 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 thrombi. In summary, the serious side effects of tamoxifen include endometrial cancer (discussed below), venous thrombosis, and cataracts.

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 one 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.\(^\text{234}\)

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.\(^\text{235}\) 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 have been reported to have less coronary heart disease in some studies, but not all.\(^\text{234, 235}\) 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 currently used.\(^\text{236}\) Tamoxifen causes a decrease in anti-thrombin III, and there has been a small increase in the incidence of thrombo-embolism observed in tamoxifen-treated patients compared with controls.\(^\text{235, 237}\) However, in the world overview of randomized trials, no significant cardiac or vascular increase in mortality was noted in tamoxifen-treated women.\(^\text{238}\)

Tamoxifen Breast Cancer Prevention Trials

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.\(^\text{238}\) This outcome was 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. In addition to these side effects, there continues to be concern that, because malignant breast tumors require a relatively long period of time to progress from an abnormal cell to a clinically detectable mass, the impact of tamoxifen may reflect growth deceleration of a pre-existing tumor. A major argument against an effect on pre-existing tumors is the documentation of persistent protection against recurrent disease over 10 years of follow-up in the overview of tamoxifen adjuvant therapy.\(^\text{239}\)

Because of the early termination of the American prevention trial, the investigators in tamoxifen prevention trials in Italy and the United Kingdom presented preliminary analyses of their results.\(^\text{240, 241}\) Both trials could detect no difference in the frequency of breast cancer between the tamoxifen and placebo arms. A consideration of power calculations indicated that despite the smaller size of these studies it should have been possible to demonstrate an effect comparable to the American trial. A prospective randomized trial of tamoxifen versus placebo in postmenopausal women was attempted in the U.K. trials contained an insufficient number of cases; thus, the impact of tamoxifen in the American trial was confined to estrogen positive cancers in older women. Another possibility, and one that is very worrisome, is that longer durations of treatment as in the European trials could allow the emergence of tamoxifen-resistant, more aggressive tumors (as has been occasionally observed with aromatase inhibitors). There is concern that tamoxifen can act as a promoter (acceleration) and anti-promoter (deceleration) effects on pre-existing tumors, not causing and preventing breast cancer. To answer this, the clinical trials must continue, with long follow-up, especially to document breast cancer mortality in these groups of women. For this reason, the Italian and U.K. investigators both decided to continue their trials. In addition, there is a fourth breast cancer prevention trial, the International Breast Cancer Intervention Study. In view of the disagreement among the American, Italian, and U.K. trials, the International Study decided to keep their study blinded, and to continue as planned.

These results lead us to recommend tamoxifen prophylaxis (20 mg daily for 5 years) for those women who are diagnosed with ductal carcinomas-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). For others who seek tamoxifen 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.

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. Toremifene has a structure and properties essentially similar to tamoxifen, and we expect the same results, side effects, and problems will be observed when this agent is used in a dose of 60 mg/day to treat breast cancer patients.

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.\(^\text{242}\) 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, 4 patients died of advanced endometrial cancer. This report from the NSABP finally convinced the breast cancer world that the risk of endometrial cancer is real with tamoxifen treatment. Although the number of deaths from endometrial cancer was small, it was the presence of 4 deaths that had a major impact. It is worth noting that the incidence of endometrial cancer in the tamoxifen-treated group was estimated to be 6.3 per 1000 patients after 5 years of treatment. This incidence is very similar to what should 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 have been reported from the Stockholm tamoxifen trial, and an increased rate of postmenopausal endometrial cancer was confirmed in the U.S. Breast Cancer Prevention Trial.\(^\text{238, 243}\) In the world overview of randomized trials, the incidence of endometrial cancer quadrupled with 5 years of tamoxifen treatment.\(^\text{244}\)

The development of endometrial cancer in women receiving tamoxifen should not be so surprising. We know that duration of exposure to estrogen is more important than the dose of estrogen in influencing progression from proliferative endometrium through hyperplasia to cancer. In addition, women being treated with tamoxifen have been reported to develop atypical hyperplasia of the endometrium, endometrial polyps, ovarian cysts, growth of fibroids, adenomyosis, and rapid exacerbation of adenomyosis.\(^\text{245, 246}\) 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.\(^\text{235}\) 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 currently used.\(^\text{236}\) Tamoxifen causes a decrease in anti-thrombin III, and there has been a small increase in the incidence of thrombo-embolism observed in tamoxifen-treated patients compared with controls.\(^\text{235, 237}\) However, in the world overview of randomized trials, no significant cardiac or vascular increase in mortality was noted in tamoxifen-treated women.\(^\text{238}\)

It is inappropriate to advocate progestational treatment to prevent the endometrial response to tamoxifen. The progesterational 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.\(^\text{247}\) Periodic endometrial aspiration biopsy, of course, would be sufficient, but this procedure carries with it the potential for a very significant negative effect on patient compliance (with her tamoxifen and with her clinician). It is argued that endometrial assessment should be limited to tamoxifen-treated women who report vaginal bleeding.\(^\text{248}\) However, in the U.K. tamoxifen prevention trial, a greater endometrial response to tamoxifen was observed in those women who developed amenorrhea.\(^\text{249}\) 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. It makes sense to detect abnormal changes as early as possible. The progestin challenge test (discussed in Chapter 18) would be a cost-effective method to detect the presence of stimulated endometrium, and a pilot study has documented its use in tamoxifen-treated women.\(^\text{247}\) However, until data are available documenting the reliability of this approach, we favor the use of ultrasonographic measurement of endometrial thickness, with saline instillation when the appearance is not totally benign (also discussed in Chapter 18). Tamoxifen is associated with an ultrasonographic image that is unique, characterized by sonolucent changes that are subendometrial, thus the usefulness of saline instillation.

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.\(^\text{250}\) In our view, an annual pelvic examination is not sufficient; every 6 months is best.

The problem of hot flushing should not be underrated. About 25% of women have vasomotor symptoms on tamoxifen, and those that already had flushing have worse flushing.\(^\text{251}\) Alternative treatments are available.

We recommend transdermal clonidine, applied with the 100 mg dose once weekly.\(^\text{252, 253}\) Side effects are minimal, and a modest impact can be expected.

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. Bellergal (a combination of belladonna alkaloids, ergotamine tartrate, and phenobarbital) treatment is slightly better than a placebo, but it is also a potent sedative.\(^\text{254}\) Veralipride, a dopamine antagonist that is active in the hypothalamus, is relatively effective in inhibiting flushing at a dose of 100 mg daily.\(^\text{255}\) Mastodynia and...
galactorrhea are the major side effects, and it should be noted that the effect of increased prolactin secretion on the risk of recurrent breast cancer is unknown. Medroxyprogesterone acetate (10–20 mg daily) and megestrol acetate (20 mg bid) are also effective, but concerns regarding exogenous steroids (especially in patients who have had breast cancer) would apply to progestins as well. 245 246 Methyldopa, in doses of 500–1000 mg/day, is approximately twice as effective as a placebo, suggesting a role for adrenoreceptors in the hot flush mechanism. 247 Venlafaxine hydrochloride is an antidepressant that inhibits serotonin re-uptake; it effectively reduced hot flush frequency in a dose of 25 mg daily. 248 Vitamin E, 800 IU daily, is only slightly more effective than placebo. 249

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 long-term tamoxifen treatment:

**All women:** Careful pelvic examination every 6 months to detect the emergence of endometriosis, ovarian cysts, uterine leiomyomata.

**Postmenopausal women:** Annual measurement of endometrial thickness by transvaginal ultrasonography. Endometrial biopsy of all women with a two-layer thickness of 5 mm or greater. Saline instillation (sonohysterography) when appearance is not totally benign.

**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 thickness measurements by ultrasonography is uncertain in premenopausal women.

Consider the use of the progesterin-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). 245 246 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.

### Needle Aspiration

Needle aspiration of breast lumps should be part of the practice of everyone who cares for women. 247 The technique is easy. A small infiltrate of xylocaine 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 onto 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. 250 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. 252 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. 253 Furthermore, the average size at which a tumor is detected (70–75% of tumors are found by patients themselves) has been (prior to mammography) 2.5 cm, a se which has a 50% incidence of lymph node involvement. To decrease the mortality from breast cancer, we must utilize a technique to find the tumors when they are smaller. Mammography is the answer. Studies of breast self-examination have been disappointing in their failure to demonstrate an impact on breast cancer stage of disease and mortality. 254

Mammography is the technique of choice. Thermography has a high rate of false positive findings, and, at best, should be considered experimental. Ultrasound can rarely reveal malignant lesions smaller than 1 cm in size. It is useful, however, to guide the aspiration of lesions. CT scanning has 2 serious limitations. The x-ray dose is large, and the slices are too thick to detect early lesions. Finally, magnetic resonance imaging is not practical because of the expense and the long scan times that are necessary.

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.
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. 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.

A pattern of dysplasia on the mammogram carries with it an increased risk (2.0–3.5 times normal) of breast cancer. The risk is similar to that seen with the other known risk factors.

The Effectiveness of Mammography

Mammography can reduce breast cancer mortality. In Nijmegen, the breast cancer mortality rate in women 35 and over was reduced by 50% by annual mammographic screening. In Utrecht, the relative risk of dying from breast cancer among screened women was reduced by 70%. The first randomized, controlled trials of mammography were begun in Sweden in 1977. Results have demonstrated a 24% reduction in mortality, 29% in women aged 50–69. Most impressively, the results in Sweden were obtained with a screening only every 2–3 years and with only a single mediolateral oblique view. Overall, a 30–40% reduction in mortality can be expected with screening mammography of asymptomatic women over age 50. A greater impact is to be expected with an increased and regular frequency of use combined with state-of-the-art technology.

About 19% more breast cancers occur in women 40–49 than in women aged 50–59, accounting for approximately 20% of all deaths due to breast cancer. But, it has been questioned whether mammography screening is effective for women under 50. The American Breast Cancer Detection Demonstration Project demonstrated that screening was just as effective for women in their 40s as in women over 50. His 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. In a randomized trial in Gothenburg, Sweden, women aged 39–49 undergoing mammographic screening every 18 months had a 45% reduction in breast cancer mortality. A meta-analysis of 7 randomized clinical trials concluded that in women aged 40–49 offered mammography screening, there was a 24% reduction in breast cancer mortality.

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 two 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. However, cancers that are detected by mammography screening 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. 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.

Younger women without risk factors must understand that approximately 50 out of every 1000 mammograms will require further diagnostic procedures, and the yield will be one invasive cancer and one non-invasive tumor. Although false-positive results are more common among younger women, the difference is not so dramatic that the overall effectiveness of screening is impaired.

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. 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. With the so-called triple approach (examination, mammography or possibly ultrasonography in young women, needle aspiration), the failure to detect a malignancy with at least one of the 3 diagnostic tests is very reliable; open biopsy can be avoided.

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.

Postmenopausal women receiving hormone therapy characteristically develop an increase in mammographic density that is often associated with breast tenderness. This increase can be focal, multifocal, or diffuse. The increase in density may be greater with continuous, combined estrogen-progestin regimens. At the present time, these findings have no prognostic or clinical significance other than making it more difficult to interpret mammograms.

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 physician beliefs. The effectiveness of mammography for women over age 75 has not been established; however, decision analysis of available data predicts a major benefit for elderly women as well. Older women need to be reminded that risk continues to increase with increasing age.

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 mammography 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.

Chapter References

Oxytocin is required for nursing but is not essential for parturition or reproductive behavior, Ishizuka B, Quigley ME, Yen SSC, Haartsen JE, Heineman MJ, Elings M, Evers JLH, Lancranjan I.


Estradiol induces a shift in cultured cells that release prolactin or growth hormone, Pulsatile GnRH stimulates normal cyclic ovarian function in amenorrheic lactating postpartum women, Double-blind controlled trial of tamoxifen therapy for mastalgia, , 1998.


Menopause and the Perimenopausal Transition

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.

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. And 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.

The Massachusetts Women's Health Study, a large and comprehensive prospective, longitudinal study of middle-aged women, provides a powerful argument that the clinical symptoms of menopause are caused by many factors other than estrogen deficiency. Women who have had the menopause for several years do not report major psychological symptoms or that their life cycles are altered. It is important, 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.
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, one of the greatest achievements of the 20th century. 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 BC, life expectancy was only 18 years. By 100 BC, the time of Caesar, it had reached 25 years. In 1900, in the United States, life expectancy still had reached only 49 years. In 2000, the average life expectancy will be 79.7 years for women and 72.9 for men. Today, once you reach 65, if you are a man you can expect to reach 80.5, if you are a woman, age 84.3. 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. 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 United States. There are more old people (with their greater needs) than ever before. In 1900, there were approximately 3 million Americans 65 and older (about 4% of the total population). By 2030, the elderly population will reach about 57 million (17% of the total population). Soon population aging will replace population growth as the most important social problem.

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 (270 million) than the number of people of all ages currently living in the U.S.

Current World Population Changes

<table>
<thead>
<tr>
<th>Year</th>
<th>Births</th>
<th>Deaths</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>140,773,000</td>
<td>51,315,000</td>
<td>89,458,000</td>
</tr>
<tr>
<td>Month</td>
<td>11,731,080</td>
<td>4,276,250</td>
<td>7,454,834</td>
</tr>
<tr>
<td>Week</td>
<td>2,707,173</td>
<td>140,589</td>
<td>245,090</td>
</tr>
<tr>
<td>Hour</td>
<td>16,070</td>
<td>5,858</td>
<td>10,212</td>
</tr>
<tr>
<td>Minute</td>
<td>268</td>
<td>96</td>
<td>170</td>
</tr>
<tr>
<td>Second</td>
<td>4.5</td>
<td>1.6</td>
<td>2.8</td>
</tr>
</tbody>
</table>

This is a worldwide development, not limited to affluent societies. 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. 95% of this growth will occur in developing countries. The poorest countries today (located in Africa and Asia) will in 2000 account 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%. Thus, it is appropriate to focus attention on population control; however, even in developing countries this will change. In 1950, only 40% of people 60 and older lived in developing countries. By 2025, more than 70% will live in those countries.
In 1900, older men in the U.S. outnumbered women 102 to 100. In the 1980s, there were only 68 men for every 100 women over the age 65. By age 85, only 40 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. Approximately 55% of girls, but only 35% of boys, live long enough to celebrate their 85th birthday.

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.

Therefore, the use of postmenopausal estrogen therapy, with its protective effect on atherosclerosis, may, in fact, exaggerate the sex differential in mortality. 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 in men: low cholesterol diet, no smoking, optimal body weight, and active exercise.

The death rate is higher for men at all ages. 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 in mortality 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.
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. Our best information comes from two 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 than 25,000 woman-years in a little over 2700 women. The observations of Vollman and Treloar documented a normal evolution in length and variation in menstrual cycles.

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 is used to define and establish what is called the perimenopausal transition. 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 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.

In the year 2000, approximately 30% of people in the U.S. who have not obtained a high school diploma will be smokers, but less than 10% of those with higher education will be smoking. Currently, approximately 28% of men and 23% of women are smokers. Cigarette smoking, therefore, continues to be the single most preventable cause of premature death in the U.S. In addition, the use of chewing tobacco, pipe smoking, and cigars contributes significantly to morbidity and mortality.

Physicians and older patients may be skeptical that quitting smoking after decades of smoking could be beneficial. In a longitudinal study of 2674 people, aged 65–74, the mortality rates for exsmokers were no higher than for nonsmokers. The effects are at least partly reversible within one to five years after quitting. Even older patients who already have coronary artery disease have improved survival if they quit smoking. 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.

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%.

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 the 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.

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.
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. 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. Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.

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. This period of longer cycles uniformly precedes menopause no matter the age when menses cease, whether menopause is early or late. The duration of the follicular phase is the major determinant of cycle length. 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 lutenizing hormone (LH) and slightly elevated levels of estradiol.

Contrary to older belief (based on the report by Sherman, et al, estradiol levels do not gradually wane in the years before the menopause, but remain in the normal range, although slightly elevated, until 6 months to 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 8 women, age 46–56. More recent longitudinal studies of women as they pass through the perimenopausal transition reveal that estrogen levels do not begin to decline until less than a year before menopause. Indeed, women experiencing the perimenopausal transition actually have higher overall estrogen levels, a response that is logically explained by an increased ovarian follicular response to the increase in FSH secretion during these years.

As noted, most women experience a 2- to 8-year period of time prior to menopause when anovulation becomes prevalent. During this period of time ovarian follicles undergo an accelerated rate of loss until eventually the supply of follicles is finally depleted. In a study of human ovaries, the accelerated loss appeared to begin when the total number of follicles reached approximately 25,000, a number reached in normal women at an age of 37–38. This loss correlates with a subtle but real increase in FSH and decrease in inhibin. The accelerated loss is probably secondary to the increase in FSH stimulation. These changes, including the increase in FSH, reflect the reduced quality and capability 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. Both inhibin A and inhibin B may be involved. Luteal phase levels of inhibin A and follicular phase levels of inhibin B decrease with aging, and may antedate the rise in FSH.

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. Thus, the changes in the later reproductive years (the decline in inhibin allowing a rise in FSH) reflect diminishing follicular reactivity and competence as the ovary ages. 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 26). 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 titrate estrogen dosage.

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. However, even under these circumstances, fluctuations can occur, with a period of ovarian failure followed by resumption of ovarian function. 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.

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. 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 postreproductive status was, for most women, approximately 4 years in duration. In the study by Treolar, 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. The mean duration of the perimenopausal transition was 5.0 years, with a range of 2 to 8 years.

The Perimenopausal Transition

Average age of onset — 46

Age of onset for 95% of women — 39 to 51
Preventive Health Screening of Healthy Perimenopausal Women

The most important thing a clinician can offer 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.
- 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.

A complete medical history and physical examination should be performed every 5 years, at about age 40, 45, 50, and 55. Annual visits should include a breast and pelvic examination, Pap test, screening for sexually transmitted diseases when appropriate, a TSH assessment in the 40s and every 2 years beginning at age 60, and hemoccult testing after age 50. Annual screening mammography should begin at age 40 (discussed in Chapter 16). At each visit, appropriate testing is scheduled for specific chronic conditions, indicated immunizations are proved, 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.

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. 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.

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 Tredlar, the average age of menopause was 50.7, and the range that included 95% of the women was 44 to 56. In a survey in the Netherlands, the average age of menopause was 50.2. About 1% of women will experience menopause before the age of 40.

Clinical impression has suggested that mothers and daughters tend to experience menopause at the same age, and there are two studies indicating that daughters of mothers with an early menopause (before age 46) also have an early menopause. There is sufficient evidence to believe that undernourished women and vegetarians experience an earlier menopause. Because of the contribution of body fat to estrogen production, thinner women experience a slightly earlier menopause. Consumption of alcohol is associated with a later menopause. This is consistent with the reports that women who consume alcohol have higher blood and urinary levels of estrogen, and greater bone density.

There is no correlation between age of menarche and age of menopause. In most studies, race, parity, and height have no influence on the age of menopause; however, two cross-sectional studies found later menopause to be associated with increasing parity. Two studies have found that irregular messes among women in their early 40s predicts an earlier menopause. A French survey detected no influence of heavy physical work on early menopause (before age 45). An earlier menopause is associated with living at high altitudes. There is reason to believe 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. And finally, earlier menopause is associated with growth retardation in late gestation.

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. Even former smokers show evidence of an impact.

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. 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. 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. 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. In the PEPI trial, 60% of women 55–64 years old were sexually active.

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. The single most significant determinant of sexual activity for older women, therefore, is the availability 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. Longitudinal studies indicate that the level of sexual activity is more stable over time than previously suggested. Individuals who are sexually active earlier in life continue to be sexually active into old age.

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. The dyspareunia associated with postmenopausal urethral 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. Less vaginal atrophy is noted in sexually active women than in inactive women; presumably the activity maintains vaginal vasculature and circulation.
Illness and Sex

It is not uncommon to encounter women who have had surgery that affects sexuality. The list includes vulvectomY, coronary bypass surgery, 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.

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 treatment. 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.

Hormone Production After Menopause

 Shortly after the menopause, one can safely say that there are no remaining ovarian follicles. 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. 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 minutes for LH and 3–4 hours for FSH), and perhaps because there is no specific negative feedback peptide for LH like inhibin.

The postmenopausal ovary secretes primarily androstenedione and testosterone. After menopause, the circulating level of androstenedione is about one half that seen prior to menopause. Most of this postmenopausal androstenedione is derived from the adrenal gland, with only a small amount secreted from the ovary, although androstenedione is the principal steroid secreted by the postmenopausal ovary. 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.

Testosterone production decreases by approximately 25% after menopause, but the postmenopausal ovary in most women, but not all, secretes more testosterone than the premenopausal ovary. With the disappearance of follicles and estrogen, the elevated gonadotropins drive the remaining stromal tissue in the ovary to a level of increased testosterone secretion. 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. 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. 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. Nevertheless, compared with young women, the overall androgen exposure of perimenopausal women to androgens is less.

<table>
<thead>
<tr>
<th>Blood Production Rates of Steroids</th>
<th>Reproductive age</th>
<th>Postmenopausal</th>
<th>Oophorectomized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androstenedione</td>
<td>2–3 mg/day</td>
<td>0.5–1.5 mg/day</td>
<td>0.4–1.2 mg/day</td>
</tr>
<tr>
<td>Dehydroepiandrosterone</td>
<td>6–9</td>
<td>1.5–4.0</td>
<td>1.5–4.0</td>
</tr>
<tr>
<td>Dehydroepiandrosterone sulfate</td>
<td>8–16</td>
<td>4–9</td>
<td>4–9</td>
</tr>
<tr>
<td>Testosterone</td>
<td>0.2–0.25</td>
<td>0.05–0.18</td>
<td>0.02–0.12</td>
</tr>
<tr>
<td>Estrogen</td>
<td>0.350</td>
<td>0.048</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Changes in Circulating Hormone Levels at Menopause

<table>
<thead>
<tr>
<th></th>
<th>Premenopause</th>
<th>Postmenopause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estradiol</td>
<td>40–400 pg/mL</td>
<td>10–20 pg/mL</td>
</tr>
<tr>
<td>Estrone</td>
<td>30–200 pg/mL</td>
<td>30–70 pg/mL</td>
</tr>
<tr>
<td>Testosterone</td>
<td>20–80 ng/dL</td>
<td>15–70 ng/dL</td>
</tr>
<tr>
<td>Androstenedione</td>
<td>60–300 ng/dL</td>
<td>30–150 ng/dL</td>
</tr>
</tbody>
</table>

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. The circulating level of estrone 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 hours, almost all, if not all, being estrogen derived from the peripheral conversion of androstenedione. 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 age, a decrease can be measured in the circulating levels of dehydroepiandrosterone sulfate (DHAS) and dehydroepiandrosterone (DHA), whereas the circulating postmenopausal levels of androstenedione, testosterone, and estrogen remain relatively constant.

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 will vary 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. Aromatization of androgens to estrogens is not limited to adipose tissue; however, because almost every tissue tested has this activity.

Eventually, the ovarian stroma 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.

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.

A precise understanding of the symptom complex the individual patient may display is often difficult to achieve. Some patients will experience severe multiple reactions that may be disabling. Others will show no reactions, or minimal reactions, that go unnoticed until careful medical evaluation.

It is helpful to classify the hormonal problems in 3 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 estrogen-progestin therapy (Chapter 19).

Problems of Estrogen Excess

Exposure to Unopposed Estrogen

Throughout the perimenopausal period, there is a significant incidence of dysfunctional uterine bleeding. 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 and nonfunctional 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 extruterine sites) must be ruled out. In addition to careful history and physical examination, dysfunctional uterine bleeding requires evaluation. Transvaginal ultrasonographic measurement of endometrial thickness can be utilized in postmenopausal women to avoid unnecessary biopsies. A total endometrial thickness greater than 4 mm is suspicious for endometrial cancer 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 traditional, more costly and risky, in-hospital dilation and curettage (D and C).

We recommend the use of the plastic endometrial suction device. It is easy to use, requires no cervical dilatation, 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 and 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 would be helpful to use such an agent at least 20 minutes 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, an in-hospital dilatation and curettage (D and C) is in order. Furthermore, if the uterus is not normal on pelvic examination, the office endometrial biopsy must yield to an in-hospital D and C with hysteroscopy in order to achieve accuracy of diagnosis.

If 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 only about 1–2% of postmenopausal endometrial biopsies. 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 will subsequently develop significant pathology within 2 years.
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).

**Hysteroscopy, hysteroscopy, or ultrasonography** with the uterine distension of saline if bleeding persists to determine the presence of endometrial polyps or submucosal fibroids.

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. The clinician must be 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 perimenopause a 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 given daily for at least the first 10 days of each month. If hyperplasia is present, follow-up aspiration curettage after 3–4 months is required, and if progestin is ineffective and histological regression is not observed, formal curettage is an essential preliminary to alternate therapeutic surgical choices. Because hyperplasia with atypia carries with it a risk of cancer (even invasive), hysterectomy is the treatment of choice.

When monthly progestin therapy reverses 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, the combined estrogen-progestin program can be initiated as needed to control the flushes.

If contraceptive use is considered, the healthy, nonsmoking patient should seriously consider the use of oral 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 vascular disease.

The flush can occur in the premenopause, but it is a major feature of the postmenopause, lasting in most women for 1–2 years but, in some (as many as 25%) for longer than 5 years. In cross-sectional surveys, up to 40% of premenopausal women and 85% of menopausal women report vasomotor complaints.

Although the flush can occur in the premenopause, it is a major feature of postmenopause, lasting in most women for 1–2 years but, in some (as many as 25%) for longer than 5 years. In cross-sectional surveys, up to 40% of premenopausal women and 85% of menopausal women report hot flushes.

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 now being focused (because of their relationship to postmenopausal hormone therapy) on cardiovascular disease and osteoporosis.

During the menopausal years, some women will experience severe multiple symptoms, whereas others will 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 biological versus cultural variability. For example, there is no word to describe a hot flush in Japanese, Chinese, and Mayan. 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, and individual perceptions.

**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” is descriptive of a sudden onset of reddening of the skin over the head, neck, and chest, accompanied by a feeling of intense body heat and concluded by sometimes profuse perspiration.

The frequency 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.

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. The frequency has been reported to be even higher in premenopausal women diagnosed with premenstrual syndrome. 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. 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 reported hot flushing.

Although the flush can occur in the premenopause, it is a major feature of postmenopause, lasting in most women for 1–2 years but, in some (as many as 25%) for longer than 5 years. In cross-sectional surveys, up to 40% of premenopausal women and 85% of menopausal women report vasomotor complaints. In the U.S., there is no difference in the prevalence of vasomotor complaints in surveys of black and white women. In a massive review of hot flushes, it was concluded that the estimates on prevalence are hampered by inconsistencies and differences in methodologies, cultures, and definitions.

The physiology of the hot flush is still not fully understood, but it appears to originate in the hypothalamus and is brought about by a decline in estrogen. However, not all hot flushes are due to estrogen deficiency. Flashes and sweating can be secondary to diseases, including pheochromocytoma, carcinoid, leukemias, pancreatic tumors, and thyroid abnormalities. 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...
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 is not borne out in objective study. 139

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. 139 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. The body surface experiences an increase in temperature, accompanied by changes in skin conductance, and 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 hypothalamic 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. 140

Premenopausal women experiencing hot flushes should be screened for thyroid disease and other illnesses. A comprehensive review of all possible causes is available. 139 Clinicians should be sensitive to the possibility of an underlying emotional problem. Looking beyond the presenting symptoms into the patient's life will be 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 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. 133

A significant clinical problem encountered in our referral practice is the following scenario: a woman will occasionally undergo an apparent 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 Women's Health and Nutrition Examination Survey indicated that estrogen use was associated with a lower prevalence of skin wrinkling and dry skin. 133 However, smoking is a major risk factor for facial skin wrinkling, and hormone therapy cannot diminish this impact of smoking. 141

A decline in skin collagen content and skin thickness that occurs with aging can be considerably avoided by postmenopausal estrogen therapy. 139, 130, 132, 134 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. 142 Although it is uncertain whether estrogen treatment can affect physical appearance, at least one study demonstrated not only an increase in facial skin thickness, but an improvement in wrinkles with topical estrogen. 132 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. 133 However, smoking is a major risk factor for facial skin wrinkling, and hormone therapy cannot diminish this impact of smoking. 141

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. However, 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. 127, 128 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. 135, 132, 130, 131 The concept of a specific 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. 136, 137, 138 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. 139 And longitudinal data from the Massachusetts Women's Health Study document that menopause is not associated with an increased risk of depression. 138 Although women are more likely to experience depression than men, this sex difference begins in early adolescence not at menopause. 132

The Hot Flush

Premenopausal: 10–25% of women

<table>
<thead>
<tr>
<th>Duration</th>
<th>No flushes</th>
<th>15–25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2 years average</td>
<td>15–20%</td>
<td>15–25%</td>
</tr>
<tr>
<td>5 plus years, 25%</td>
<td>15–20%</td>
<td>15–25%</td>
</tr>
</tbody>
</table>

Other causes:
- Psychosomatic
- Stress
- Thyroid disease
- Pheochromocytoma
- Carcinoid
- Leukemia
- Cancer

FSH.
Female survivors of a first myocardial infarction face the prospects of a second myocardial infarction in 31%, a stroke in 18%, and sudden death in 6%. Since 1984, the number of cardiovascular disease deaths for women has exceeded those for men.

Cardiovascular Disease

Long-term treatment of perimenopausal women with estrogen users. In 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. The effect was greater with increasing dose and duration of use (there was no benefit that resulted from long duration of use). There were too few patients in the Rancho Bernardo cohort with dementia to assess the impact of estrogen therapy. An improvement in insomnia with estrogen treatment can even be documented in postmenopausal women who are reportedly asymptomatic.

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 an indirect benefit of relief from physical symptoms and, consequently, improved sleep. Utilizing various assessment tools for measuring depression, improvements with estrogen treatment have been recorded in oophorectomized women. 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. Indeed, treated women had higher depressive symptom scores, presumably reflecting treatment selection bias: symptomatic and depressed women seek hormone therapy. Nevertheless, estrogen therapy is reported to have a more powerful impact on women’s well-being beyond the relief of symptoms such as hot flushes. In elderly depressed women, improvements in response to fluoxetine were enhanced by the addition of estrogen therapy.

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. However, the effects on 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. Another short-term (3 months) randomized, double-blind study could detect no improvement in cognitive performance compared with placebo treatment. 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 women postmenopausal women taking estrogen scored higher on tests of immediate and delayed recall. 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. Women in the Baltimore Longitudinal Study of Aging who were using estrogen performed better in a test of short-term visual memory. 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. 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 progesterational agents. Fatigue, irritability, headache, and depression are not thought to be estrogen-related phenomena.

Tests of cognitive function in the prospective cohort study of the Rancho Bernardo retirement community failed to support a positive impact of postmenopausal estrogen on mental function in old age. The Rancho Bernardo study has the attribute of being large in size (800 women), which also allowed for testing of duration of use (there was no benefit that resulted from long duration of use). There were too few patients in the Rancho Bernardo cohort with dementia to assess the impact of estrogen on Alzheimer’s disease.

Up to 3 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. Progestational agents do not exert similar actions.

Important early findings have 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. 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. Furthermore, the administration of estrogen to patients with Alzheimer’s disease has improved cognitive performance. The presence of estrogen therapy also enhances the beneficial response to tacrine in women with Alzheimer’s disease. If this effect of estrogen can be verified in the ongoing clinical trials, the potential impact on this dreadful condition that is more prevalent in women cannot be underrated. This would be another argument in favor of long-term treatment.

Cardiovascular Disease

Diseases of the heart are the leading cause of death for women in the United States, followed by malignant neoplasms, cerebrovascular disease, and motor vehicle accidents. Since 1984, the number of cardiovascular disease deaths for women has exceeded those for men. About 20,000 women under age 65 die of myocardial infarction each year in the U.S. The death rate for coronary heart disease in women is approximately 3 times greater than the rates for breast cancer and lung cancer. Female survivors of a first myocardial infarction face the prospects of a second myocardial infarction in 31%, a stroke in 18%, and sudden death in 6%.
Most cardiovascular disease results from atherosclerosis in major vessels. The risk factors are the same for men and women: high blood pressure, smoking, diabetes mellitus, and obesity. However, when controlling for these risk factors, men have a risk of developing coronary heart disease over 3.5 times 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 LDL, producing a modified LDL that is chemotactic for circulating monocytes, that 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. The mechanism of this adaptive remodeling is not known, but the extent of this process must affect the risk of occlusion and infarction.

The fatty streak in arterial vessels is 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 process is initiated by the aggregation and adherence of circulating monocytes to a site on the arterial endothelium. When the monocytes penetrate through the endothelium and enter the intima, they become loaded with lipids and converted to the foam cells. Modification of LDL, especially oxidation, is crucial in this conversion of monocytes to foam cells. The initial step, adherence of monocytes to endothelium, can be induced by elevated cholesterol and LDL-cholesterol in the circulation.

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 had 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. 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 low-density (LDL)-cholesterol levels are lower in premenopausal women than in men, although the levels gradually increase with aging and after menopause they rise rapidly. After menopause the risk of coronary heart disease doubles for women as the atherogenic lipids about age 60 reach levels greater than those in men. These changes can be favorably reduced by dietary changes. Of course, these lipid changes at menopause (whether natural or surgical) can be reversed with estrogen treatment. At all ages, however, HDL-cholesterol values in women are 10 mg/dL higher than in men.

Prospective studies have documented the strong association between total cholesterol and coronary heart disease in women; however, coronary heart disease risk appears at higher total cholesterol levels for women than for men. 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. This is the reason for ceasing lipoprotein screening after age 75 in patients with normal lipids. However, this decision should be individualized, taking
The strongest predictor of coronary heart disease in women is a low HDL-cholesterol. 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 have 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%. Women with very high HDL-cholesterol levels (over 55–60 mg/dL) have virtually no increased risk of heart disease even in the presence of elevated concentrations of total cholesterol. It is appropriate to be concerned when HDL-cholesterol levels are less than 50 mg/dL.

### The Optimal Cholesterol/Lipoprotein Profile

<table>
<thead>
<tr>
<th>Total cholesterol</th>
<th>Less than 200 mg/dL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL-cholesterol</td>
<td>Greater than 50 mg/dL</td>
</tr>
<tr>
<td>LDL-cholesterol</td>
<td>Less than 130 mg/dL</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>Less than 250 mg/dL</td>
</tr>
</tbody>
</table>

In some studies, elevated LDL-cholesterol has been identified as a predictor of coronary heart disease in women, but in others it has not. 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.

Triglycerides are also an important risk factor for coronary heart disease in women; however, an increased rate of cardiovascular disease is observed only when increased triglyceride levels are present in association with low HDL-cholesterol levels. 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 200–400 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.

The proportion of HDL-cholesterol in the total cholesterol decreases with age in women and results in an age-related increase in the ratio of total cholesterol to HDL-cholesterol, from 3.4 at ages 25–34 years to 4.7 at ages 75–89 years. When the ratio exceeds 7.5, women have the same coronary heart disease risk as men. The optimal ratio is about 3.5, and risk increases significantly above 5.0. For ease of remembering, consider anything below 4.0 as good and any ratio above 4.0 as unhealthy. However, attention should not be totally focused on this ratio; the LDL-cholesterol and triglyceride levels cannot be ignored, and the entire lipid profile should be assessed.

### Heart Disease Risk Based on Cholesterol/HDL Ratio

<table>
<thead>
<tr>
<th>Lowest risk</th>
<th>Less than 2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below average risk</td>
<td>2.5–3.7</td>
</tr>
<tr>
<td>Average risk</td>
<td>3.8–5.6</td>
</tr>
<tr>
<td>High risk</td>
<td>5.7–8.3</td>
</tr>
<tr>
<td>Dangerous</td>
<td>Greater than 8.3</td>
</tr>
</tbody>
</table>

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. Estrogen inhibits the oxidation of LDL, and also protects against the toxic effects of oxidized LDL on the endothelium.

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 hypercholesterolemia 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. The speculation is that impaired liver growth in utero produces a permanent adverse change in cholesterol and lipoprotein metabolism. Reduced fetal growth also leads to insulin resistance and lower HDL-cholesterol levels, most severe in those who become obese.

Another newly appreciated factor is the connection between hormones and fat. Adiposity of the trunk is a risk factor for coronary heart disease in women and is associated with a relatively androgenic hormonal state, hypertension, and disorders of lipid and carbohydrate metabolism. Central fat distribution in women is positively correlated with increases in total cholesterol, triglycerides, and LDL and negatively correlated with HDL. The atherogenic lipid profile associated with abdominal adiposity is at least partly mediated through an interplay with insulin and estrogen. 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. 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, but it reflects diet, exercise, and aging.

In the last 30 years, stroke mortality has declined by 60% and mortality from coronary heart disease by more than 50% 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. It is now recognized that there is a strong and growing scientific basis for preventive medicine and health promotion efforts in clinical practice.

We now have another opportunity. Postmenopausal hormone therapy deserves consideration as a legitimate component of preventive health care for older women. One can argue convincingly that protection against cardiovascular disease is the major benefit of postmenopausal estrogen treatment, and the magnitude of this benefit is considerable. There is a sound rationale for this protection in the link between cardiovascular disease and the sex hormones.

### The Evidence for Protection Against Cardiovascular Disease by Estrogen

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. A population-based case-control study in Rochester, Minnesota, concluded that if all eligible women utilized estrogen, myocardial...
Hypertension is both a risk factor for cardiovascular mortality and a common problem in older people. It is important, therefore, to know that no relationship has been documented between postmenopausal estrogen use and the risk of stroke. The Epidemiologic Follow-up Study demonstrated a 30% reduction in stroke incidence among users of estrogen, and the Nurses' Health Study showed a 31% reduction in stroke incidence and a strongly significant 63% reduction in stroke mortality. These epidemiologists concluded that their evidence strongly indicates that in women receiving estrogen therapy, the risk of cardiovascular disease is reduced by 40% to 60%, and the risk of stroke is reduced by 50% to 60%. Long-term estrogen users (10–25 years) were identified in the Kaiser Permanente Medical Care Program in California and found to have a 46% reduced risk of all-cause mortality, largely due to reductions in cardiovascular disease. Women who used estrogen for more than 15 years had a 30% greater reduction in the risk for mortality than short-term users.

An important question has been raised, asking whether estrogen treatment is a marker for variables (such as better diet and better health care) that place women at a lower risk of stroke. These epidemiologists concluded that their evidence strongly indicates that estrogen use reduces the risk of cardiovascular disease and stroke. The Nurses' Health Study cohort (a prospective study of women aged 45–64 years) demonstrated a 3% reduction in the risk of stroke among estrogen users, indicating that selection of significantly more healthy women for estrogen use cannot fully explain the beneficial effect of estrogen on the risk of cardiovascular disease. In a comparison of health variables among users and non-users in south Australia, there was no evidence to support the presence of a "healthy user" effect. However, in the Nurses' Health Study, estrogen use was associated with a 46% overall reduction in the risk of death from stroke, with a 70% reduction in recent users.

In a large Danish case-control study, no impact was detected of either estrogen or combined estrogen and progestin on the risk of non-fatal stroke, both thromboembolic and hemorrhagic. A case-control study from Seattle found a 50% reduced risk of subarachnoid hemorrhage with the use of postmenopausal hormone therapy, and the effect was even greater among smokers. Positive relationships with duration and recency of use argued in favor of a causative association. In the prospective study of the Leisure World cohort, estrogen therapy was associated with a 46% overall reduction in the risk of death from stroke, with a 70% reduction in recent users. This protection was present in both women with and without hypertension and in both smokers and nonsmokers. This level of protection was similar to that observed in the same Leisure World population for estrogen protection against deaths due to myocardial infarction. The failure to observe a protective effect in the Nurses' Health Study cohort is in striking contrast to the Leisure World cohort. Indeed, in the Nurses' Health Study cohort, there was a slight increase (not statistically significant) in risk for thromboembolic stroke in current estrogen users. The population-based cohort study in Uppsala, Sweden, documented a 30% reduced incidence of stroke in postmenopausal users of estrogen, and, importantly, women prescribed an estrogen-progestin combination, containing a significant dose of the potent androgenic agent levonorgestrel, also experienced a reduced incidence of stroke. A reduced risk for mortality from stroke in this Swedish study was confined to intracerebral hemorrhage.

Within this confusing mixture of results, there is 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 I) conducted from 1971 to 1974 demonstrated a 25% reduction in the risk of stroke in women aged 45–64 years. This reduction was significant, and in women using estrogen, the risk was reduced by 31%.

In cohort studies, only two produced conflicting data. The Walnut Creek Study, one initially with conflicting data, had, in its first report, only 26 women with myocardial infarctions, and only 9 were estrogen users. An update of the Walnut Creek data documented a 50% reduction in death from strokes in both estrogen users. The Framingham Heart Study presented data in 1978, and in 1985, which argued that there was a 50% increased risk for cardiovascular disease among estrogen users, although there was no difference in fatality rates between users and nonusers. Because of the respect the Framingham Heart Study carries, its impact was significant. Subsequent re-analysis of the Framingham heart data eliminated angina as a consideration by the authors of the study reversed their conclusion. The early reports from the Framingham Heart Study, therefore, stand in lonely opposition to overwhelming evidence that appropriately low doses of estrogen protect postmenopausal women against cardiovascular disease.

In the Nurses' Health Study with 16 years of follow-up, the age-adjusted relative risk of coronary disease in current users of estrogen alone was 40% reduced, and with estrogen and progestin, 61% reduced. No association between use and stroke was observed, but there were relatively few strokes in this cohort. Adjustments for diet, use of vitamins, use of aspirin, and physical activity did not change the conclusions. 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.

Stroke

The incidence of stroke in women aged 45–64 years (1–2 per 1000 per year) has changed little in the past few decades; however, mortality rates have decreased. 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 20 years regarding estrogen use and stroke have not been consistent. The many studies have indicated either no effect of postmenopausal hormone therapy on the risk of stroke or a reduction in risk associated with estrogen or estrogen-progestin use.

In a large Danish case-control study, no impact was detected of either estrogen or combined estrogen and progestin on the risk of non-fatal stroke, both thromboembolic and hemorrhagic. A case-control study from Seattle found a 50% reduced risk of subarachnoid hemorrhage with the use of postmenopausal hormone therapy, and the effect was even greater among smokers. Positive relationships with duration and recency of use argued in favor of a causative association. In the prospective study of the Leisure World cohort, estrogen therapy was associated with a 46% overall reduction in the risk of death from stroke, with a 70% reduction in recent users. This protection was present in both women with and without hypertension and in both smokers and nonsmokers. This level of protection was similar to that observed in the same Leisure World population for estrogen protection against deaths due to myocardial infarction. The failure to observe protection against stroke in the Nurses' Health Study cohort is in striking contrast to the Leisure World cohort. Indeed, in the Nurses' Health Study cohort, there was a slight increase (not statistically significant) in risk for thromboembolic stroke in current estrogen users. The population-based cohort study in Uppsala, Sweden, documented a 30% reduced incidence of stroke in postmenopausal users of estrogen, and, importantly, women prescribed an estrogen-progestin combination, containing a significant dose of the potent androgenic agent levonorgestrel, also experienced a reduced incidence of stroke. A reduced risk for mortality from stroke in this Swedish study was confined to intracerebral hemorrhage.

Within this confusing mixture of results, there is 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 I) conducted from 1971 to 1974 demonstrated a 25% reduction in the risk of stroke in women aged 45–64 years. This reduction was significant, and in women using estrogen, the risk was reduced by 31%.

Hypertension

Hypertension has both a risk factor for cardiovascular mortality and a common problem in older people. It is important, therefore, to know that no relationship has been established between hypertension and the doses of estrogen used for postmenopausal therapy. Studies have either shown no effect or a small, but statistically significant, decrease in blood pressure due to estrogen treatment. This has been the case in both normotensive and hypertensive women. The addition of a progestin does not affect this response. 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.
The acute administration of estrogen to women with hypertension is followed by decreases in blood pressure, pulse rate, and circulating levels of norepinephrine.

The very rare cases of increased blood pressure due to oral estrogen therapy truly represent idiosyncratic reactions. Because of the protective impact of appropriate estrogen treatment on the risk of cardiovascular disease, it can be argued that a woman with controlled hypertension is in need of that specific benefit of estrogen. Indeed, in the Nurses’ Health Study, hypertensive women who used postmenopausal hormone therapy did have a reduced risk of coronary heart disease. Blood pressure should be assessed every 6 months in hypertensive women being treated with postmenopausal hormones, and if the blood pressure is labile, every 3 months.

Mechanisms for Estrogen's Protection Against Cardiovascular Disease

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. Although it is now recognized that this is not the sole mechanism for estrogen’s beneficial impact on cardiovascular disease, the lipid effects still play a substantial role.

Estrogen increases triglyceride levels and increases LDL catabolism as well as lipoprotein receptor numbers and activity, resulting in decreasing LDL levels. The increase in HDL levels, particularly due to the HDL₃ subfraction, is an important degree the consequence of the inhibition of hepatic lipase activity, which converts LDL₃ to LDL₂. Postmenopausal estrogen therapy with or without added progesterone also produces a beneficial reduction in the circulating levels of lipoprotein(a).

LDL particle size gets smaller (potentially a more atherogenic adverse effect) with estrogen treatment, a change that is associated with an increase in the triglyceride content of LDL; however, it is not certain to what degree this change is related to dose nor is the clinical significance understood. The addition of a progestin does not affect this development of postmenopausal women. Furthermore, but not as small LDLs is more atherogenic. This is because it is more easily oxidized, the oxidized LDL particles are may be compensated by the antioxidant activity of estrogen, which protects against atherosclerosis by inhibiting lipoprotein oxidation. Indeed, in the monkey, a smaller, more dense LDL is produced in response to estrogen, and this smaller LDL is less atherogenic. In addition, the increase in triglycerides is associated with an increase in a form (buoyant) of very low density lipoprotein (VLDL), a form that is more rapidly transported and thus less likely to be atherogenic.

Therefore, estrogen induces a change in LDL 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.

The changes in circulating apoprotein levels mirror those of the lipoproteins: apolipoprotein B (the principal surface protein of LDL) levels diminish in response to estrogen, and apolipoprotein A-I (the principal apolipoprotein of HDL) increases. The HDL and triglyceride increases induced by estrogen treatment are attenuated if progestins are added in sufficient doses. Both estrogen and exercise favorably affect the lipids and lipoprotein profile; however, in a comparison of exercise and estrogen treatment, combining estrogen and exercise did not produce a synergistic outcome. On the other hand, 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.

The degree to which estrogen-induced lipid changes contribute to the overall protection against cardiovascular disease exerted by estrogen is uncertain. One analysis suggested that 25% of the estrogen effect could be attributed to lipid changes. Certainly the lipid story is not as important as we once thought it was; nevertheless, it should not be underestimated. One study concluded that estrogen-induced lipid changes could reduce the risk of coronary artery disease by about 50% in women who have abnormal lipids.

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 progesterin to monkeys fed a high-cholesterol diet decreased the extent of coronary atherosclerosis despite a reduction in HDL-cholesterol levels. In somewhat similar experiments, estrogen treatment markedly prevented arterial lesion development in rabbits, and this effect was not reduced by adding progesterin to the treatment regimen. In this same rabbit model, raloxifene also inhibited arterial lesion formation. 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 progessional agents, lowering of HDL is not necessarily atherogenic if accompanied by an increased estrogen impact.

The monkey studies have been 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. A direct inhibition of LDL accumulation and an increase in LDL metabolism in arterial vessels could be demonstrated in these monkeys being fed a highly atherogenic diet. However, the daily administration of medroxyprogesterone acetate in this monkey model prevented the beneficial effect of conjugated estrogen on coronary artery atherosclerosis.

In postmenopausal women, the prevalence of atherosclerosis is reduced either using estrogen alone or combined estrogen-progesterin therapy as documented in the coronary arteries by CT scanning and in the carotid arteries by ultrasonography.

Thus, estrogen exerts a protective effect directly on the arterial wall independent of its effects on circulating lipoproteins. The presence of sex steroid receptors in arterial endothelium and smooth muscle lends support for the importance of this direct action; however, the mechanisms involved remain unknown. One important mechanism may be coronary artery remodeling. Atherosclerosis is associated with a compensatory change in coronary artery size that increases the magnitude of the lumen (a process called remodeling). It is important to note that the addition of a progestin did not attenuate the beneficial effect of estrogen on coronary remodeling in monkeys.

Endothelium-Dependent Vasodilatation 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 (EDRFs and EDCF). In hypertension and other cardiovascular diseases, the release of EDRFs (which is probably one factor, nitric oxide) is blunted, and the release of EDCF (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.

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). These local actions are a likely site for sex steroid involvement; the vasodilating and antiplatelet action of estrogen, especially in the coronary arteries, is a consequence of endothelial responses. In addition, postmenopausal women experience greater stress-induced neuroendocrine and blood pressure responses that are ameliorated by estrogen treatment. However, the addition of medroxyprogesterone acetate (10 mg for 10 days) attenuated the beneficial effects of estrogen on measures of cardiovascular reactivity.
Increased blood flow due to vasodilatation 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. The synthesis and secretion of nitric oxide (the potent endothelial vasodilating product) can be directly stimulated by estrogen in vitro experimental preparations of coronary arteries. In both normal postmenopausal women and women with hypertension, hypercholesterolemia, diabetes mellitus, or coronary artery disease, the intra-arterial infusion of physiologic amounts of estradiol into the forearm potentiates endothelium-dependent vasodilatation, and there is a dose-response effect with enhanced responses limited to the production of high blood levels of estradiol. Nevertheless, comparing brachial artery responses in women who are long-term hormone users (with or without progestin) with non-users, improved endothelium-dependent vasodilatation could be observed with standard doses. In a careful, randomized study, the addition of norethindrone acetate did not reduce the beneficial effect of estrogen on peripheral artery blood flow. 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.

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) catalyze 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. In animal experiments, the endothelial basal release of nitric oxide is greater in females, a gender difference that is mediated by estrogen. 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. In contrast, long-term treatment with estradiol and norethindrone acetate was not associated with changes in nitric oxide, endothelin-1, prostacyclin, or thromboxane A₂.

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 vasocostriction into vasodilatation with increased flow. 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). This same estrogen-associated response is observed in women with coronary atherosclerosis comparing estrogen users to non-users. This is an endothelium-dependent response, mediated to a significant degree by an increase in nitric oxide. The administration of standard doses of estrogen to women with coronary artery disease reduces the degree of ischaemia and delays the onset of signs of myocardial ischaemia on electrocardiograms, and increases exercise tolerance.

In normal women, the standard oral 0.625 mg dose of conjugated estrogens had no effect on hemodynamic responses to treadmill exercise. In the monkey, the vasodilatory response to acetylcholine requires a blood level of estradiol higher than 60 pg/mL.

Endothelium-Independent Vasodilatation

Estrogen causes relaxation in coronary arteries that are denuded of epithelium. This response is not prevented by the presence of inhibitors of nitric oxide synthase or prostaglandin synthase. Thus this vasodilatation is achieved through a mechanism independent of the vascular endothelium, perhaps acting on calcium-mediated events. The vasodilatation 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 increases forearm vasodilatation induced by sodium nitroprusside.

Inotropic Actions on the Heart and Large Blood Vessels

Estrogen treatment increases left ventricular diastolic filling and stroke volume. This effect is probably a direct inotropic action of estrogen that delays the age-related change in compliance that impairs cardiac relaxation. 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. On the other hand, others have detected attenuation of estrogen's beneficial effects on compliance (stiffness) associated with combined estrogen-progestin treatment. 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. The reasons for these differences are not apparent.

Improvement of Glucose Metabolism

An age-related decline in the basal metabolic rate begins at menopause, associated with an increase in body fat, especially central (android) body fat. Insulin resistance and circulating insulin levels increase in women after menopause, and impaired glucose tolerance predicts an increased risk of coronary heart disease. Estrogen (with or without progestin) prevents the tendency to increase central body fat with aging. This would inhibit the interaction among abdominal adiposity, hormones, insulin resistance, hyperinsulinemia, blood pressure, and an atherogenic lipid profile. Hyperinsulinemia also has a direct atherogenic effect on blood vessels, perhaps secondary to insulin propopdides. In addition to its vasocostrictive 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.

Postmenopausal women being treated with oral estrogen have lower fasting insulin levels and a lesser insulin response to glucose, indicating another mechanism for the protection against cardiovascular disease. However, not all studies have observed these favorable changes. There is an important dose-response relationship; 0.625 mg conjugated estrogens improved insulin sensitivity, 1.25 mg did not. The addition of a progestational agent can attenuate this beneficial response to oral estrogen, but the clinical impact and variation with dose are unknown. 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. Oral administration of estrogen has little effect on insulin metabolism, unless a dose is administered that is equivalent to 1.25 mg conjugated estrogens. Because a lower oral dose produces a beneficial impact, this suggests that the hepatic first-pass effect is important in this response.

Postmenopausal women with diabetes mellitus can benefit from the cardioprotective actions of estrogen. In addition, as noted above, estrogen may improve the metabolic changes associated with diabetes. Indeed, in double-blind, cross-over, placebo-controlled studies of postmenopausal women with noninsulin-dependent diabetes mellitus, estrogen treatment improved all glucose metabolic parameters (including insulin resistance), the lipoprotein profile, and measurements of androgenicity. These changes would reduce the risk of cardiovascular disease, however, long-term studies are yet to be available.

Inhibition of Lipoprotein Oxidation
The oxidation of LDL 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. Vitamin E and beta-carotene (the prohormone of vitamin A) are antioxidants. Studies indicate that the risk of myocardial infarction in smokers is higher in men and women with the lowest carotene levels, and supplementary intake of vitamin E decreases the risk of coronary heart disease. The Nurses' Health Study reported a decreased risk of coronary heart disease in women with high carotene levels. Thus, treatment with antioxidants may reduce the risk of cardiovascular disease and the risk of complications in those who already have cardiovascular disease.

Estrogen is an antioxidant. Estradiol directly inhibits LDL oxidation in response to copper, and decreases the overall formation of lipid oxides. Importantly, this antioxidant action of estradiol is associated with physiologic blood levels. In addition, estrogen may regenerate circulating antioxidants (tocopherols and beta-carotene) and preserve these antioxidants within LDL particles. This antioxidant action of estrogen preserves endothelial-dependent vasodilator function by preventing the deleterious effect that oxidized LDL has on endothelial production of vasoactive agents. There is evidence in the monkey in a short-term study that progesterone partially attenuates the inhibition of LDL oxidation by estradiol. However, 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. And in a 1-year study, the presence of levonorgestrel did not attenuate the antioxidant activity of estradiol.

A Favorable Impact on Fibrinolysis

Menopause is followed by increases in factor VII, fibrinogen, and plasminogen activator inhibitor-1 (PAI-1). These changes produce a relatively hypercoagulable 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 plasminogen activator inhibitor-1 (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 menoopause. This would be consistent with increased fibrinolytic activity, another 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. However, in a randomized 1-year trial, the addition of medroxyprogesterone acetate to a sequential regimen had greater antioxidant activity compared with the days on estrogen alone.

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. In regards to PAI-1, studies with transdermal estrogen have provided us with conflicting data; for example, favorable changes in PAI-1 levels as well as no effect. 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. Appropriate doses of hormone therapy have been reported to not have an adverse impact on clotting factors. However, one study found slightly increased clotting activation with transdermal administration of estradiol, but no change with oral conjugated estrogens. 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 one and an increase in another. The clotting story is difficult to unravel. Perhaps one contributor to the uncertainty is a possible difference between short-term and long-term effects.

Coagulation and Fibrinolysis Factors

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
  - Plasminogen activator inhibitor-1 (PAI-1)
- Factors that favor clotting when decreased
  - Antiplasmin

Overall, estrogen treatment is associated with favorable changes consistent with an increase in fibrinolysis. The favorable changes in the clotting system may contribute to estrogen’s protection of the cardiovascular system by inhibiting arterial thrombosis. How can there be a beneficial effect on arterial thrombosis when there is an increased risk of venous thrombosis (discussed in Chapter 18)? 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. Decreases in antithrombin III and protein S associated with estrogen treatment, a hypercoagulable change, may have a greater impact on the venous system.

It is important to recognize the importance of the dose-response relationship between estrogen and the risk of arterial thrombosis. We learned from our experience with high-dose oral contraceptives that high doses of estrogen cause arterial thrombosis. Low doses of estrogen may even protect against arterial thrombosis. Thus, we refer to this important dose effect as a "narrow therapeutic window." (also discussed in Chapter 18) Doses of estrogen greater than those that protect against cardiovascular disease and osteoporosis (equivalent to 0.625 mg conjugated estrogens) should be avoided. At least 3 studies have found that estrogen doses greater than 0.625 mg conjugated estrogens are less beneficial in terms of coronary heart disease and mortality; however, the numbers of patients on the higher doses were relatively small, and these conclusions did not achieve statistical significance.

Inhibition of Vascular Smooth Muscle Growth and Migration — 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. Nitric oxide, which is regulated by estrogen, also inhibits smooth muscle proliferation and migration. This smooth muscle response occurs following vascular injury, and this injury-induced reaction is inhibited by estradiol even in a mouse model deficient in the estrogen receptor-alpha, suggesting that the estrogen receptor-beta is an
Reduced levels of angiotensin-converting enzyme and renin.

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. The decrease in intimal thickening and recovery of the endothelium in response to injury 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. In the rat, medroxyprogesterone acetate blocked the estrogen-induced healing response after carotid artery injuries. 287


Protection of Endothelial Cells

Endothelial cells can respond to injury by initiating the clotting process or by atherosclerosis. Animals studies indicate that estrogen accelerates healing and recovery of the endothelium in response to injury. 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. In the rat, medroxyprogesterone acetate blocked the estrogen-induced healing response after carotid artery injuries. 287

Inhibition of Macrophage Foam Cell Formation

A feature of atherosclerotic plaque formation is monocyte infiltration into the arterial wall and the formation of macropage foam cells. In a non-antioxidant activity, estrogen inhibits macropage foam cell accumulation in atherosclerotic lesions. 288

Reduction of Angiotensin-Converting Enzyme (ACE) and Renin Levels

Although oral estrogen, but not transdermal estrogen, increases angiotensinogen levels, ACE and renin levels are decreased (with or without progestin) by both routes of administration. 289 290

Reduction of P-Selectin Levels

P-selectin is an adhesion glycoprotein that recruits leukocytes to the endothelium and plays a role in attaching platelets to endothelium. It is stored in and mobilized from platelets. Increased P-selectin expression has been demonstrated in atherosclerotic endothelial cells, and premenopausal women have lower circulating levels compared with men. Estrogen treatment lowers circulating levels of P-selectin, a change that should contribute to protection against atherosclerosis. 291

Reduction of Homocysteine Levels

Increased circulating levels of homocysteine are correlated with increased risks of atherosclerosis and thrombosis. Homocysteine levels increase after menopause, and these levels are significantly lowered by estrogen or estrogen-progesterin treatment. 292

H3>Summary — Estrogen and the Cardiovascular System

The possible beneficial actions of estrogens on cardiovascular disease include all of the following. A study in mice lacking the estrogen receptor-alpha suggests that some (perhaps many or all) of these actions are mediated by the estrogen receptor-beta or by nongenomic actions of estrogen. ER-a and ER-b are both prevalent in coronary and aortic smooth muscle cells from monkeys. Only ER-b is expressed in umbilical vein endothelial cells. 293

1. A favorable impact on the circulating lipid and lipoprotein profile, especially a decrease in total cholesterol and LDL-cholesterol and an increase in HDL-cholesterol.
2. A direct antatherosclerotic effect in arteries.
3. Augmentation of vasodilating and antiplatelet aggregation factors, specifically nitric oxide and prostacyclin (endothelium-dependent mechanisms).
5. Direct inotropic actions on the heart and large blood vessels.
6. Improvement of peripheral glucose metabolism with a subsequent decrease in circulating insulin levels.
7. Antioxidant activity.
8. Favorable impact on fibrinolysis, at least partially mediated by endothelial nitric oxide and prostacyclin synthesis.
9. Inhibition of vascular smooth muscle cell growth and migration — intimal thickening.
10. Protection of endothelial cells from injury.
11. Inhibition of macrophage foam cell formation.
12. Reduced levels of angiotensin-converting enzyme and renin.
13. Reduction of P-selectin levels.
14. Reduction of homocysteine levels.

Cardiovascular Disease and Progestins

Because the public health benefit of estrogen therapy on cardiovascular disease is of such enormous impact, it is vital that we know whether the addition of progestin attenuates estrogen’s effects on the cardiovascular system. Conclusions regarding the impact of progestational agents on cardiovascular disease are very much influenced by dose and duration of administration of the progestational agent involved. While short-term studies suggest a negative impact of progestin (i.e., subtracting from the beneficial effect of estrogen), long-term studies indicate that this short-term effect disappears.

In sequential regimens, a review of the literature on this issue suggests a dose-response relationship. A decrease in HDL-cholesterol has been noted with 10-day monthly treatment with norethindrone (5 mg), megestrol acetate (5 mg), levonorgestrel (250 µg), and even medroxyprogesterone acetate (10 mg). No significant change was noted with micronized progesterone (200 mg). The lack of an effect noted with micronized progesterone was observed with a dose (200 mg daily) that yields a normal luteal phase blood level of progesterone. A similar “physiologic” dose of synthetic progestins may be free of an adverse impact on HDL-cholesterol. Barrett-Connor and colleagues have been studying the adult residents of Rancho Bernado, California. The women using both estrogen and progestin demonstrated the same favorable impact on cardiovascular risk factors as estrogen-only users compared with the nonusers. On the other hand, a well-designed study indicated that although the sequential estrogen plus progestin program had a favorable impact on lipids and the lipoprotein profile, the impact (as measured one year later) was less than that achieved by estrogen alone. 294

Studies with the combination of an estrogen and a low dose of a progestational agent administered continuously (every day without a break) are documenting a favorable impact on total cholesterol and LDL, with attenuation of the estrogen-induced increase in HDL. The various formulations include estradiol and levonorgestrel, 295 estradiol and 1 mg norethindrone acetate, 296 estradiol and dydrogesterone, 297 estradiol valerate and levonorgestrel or lynprogestrone acetate, 298 ethinyl estradiol and 0.5–1.0 mg norethindrone acetate, 299 300 and 0.625 conjugated estrogen and 2.5–5.0 mg medroxyprogesterone acetate. 301 Christiansen has documented the maintenance of a favorable lipid profile over a period of 5 years of treatment with continuous, combined estradiol and 1 mg norethindrone acetate. 302

A large, prospective, 1-year randomized clinical trial compared sequential and combined regimens of conjugated estrogens and medroxyprogesterone acetate with unopposed estrogen. Although the increase in HDL-cholesterol, HDL- and apoliprotein A-I levels was greater with unopposed estrogen, there was still a significant increase with combined estrogen-progestin treatment. Importantly, the decrease in total cholesterol, LDL-cholesterol, apolipoprotein B, and Lp(a) was equivalent, comparing unopposed estrogen with either sequential or daily, combined estrogen-progestin. Fasting glucose and insulin levels were improved in all treatment groups. Similar results have been observed in a 2-year randomized trial with 2 mg estradiol and 1 mg norethindrone acetate. A prospective case-control study indicated that the decrease in Lp(a) achieved by estrogen is attenuated by the addition of a progestin. In a large cross-sectional study of 4958 postmenopausal women, a favorable impact on the lipoprotein profile and fasting insulin levels was observed in users of unopposed estrogen and combined estrogen-progestin compared with nonusers. In Spain, 8 months of uninterrupted treatment with a daily dose of 2.5 mg medroxyprogesterone acetate in a combined estrogen-progestin program produced a favorable effect on the lipoprotein profile not substantially different from that achieved with estrogen alone. Nevertheless, the bulk of evidence indicates that the addition of a progestin attenuates, but does not eliminate, the favorable increase in HDL achieved with estrogen treatment.
whereas the beneficial decreases in total cholesterol, LDL, lipoprotein(a), and the apolipoproteins are maintained. 424

Studies in women of arterial vascular resistance by Doppler ultrasound flow patterns have indicated no detrimental effects of exogenous progesterone or medroxyprogesterone acetate to the beneficial decrease in resistance (and increase in blood flow) produced by estrogen administration. 425 426

The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial

PEPI was a randomized, double-blind, placebo-controlled, 3-year trial of 875 postmenopausal women conducted in 7 clinical centers in the U.S. 422 427 The participants were randomly assigned to the following treatment groups:

1. Placebo.
2. Estrogen alone, 0.625 mg conjugated estrogens daily.
3. Sequential estrogen and progesterone, 0.625 mg conjugated estrogens daily and 10 mg medroxyprogesterone acetate days 1–12.
4. Sequential estrogen and progesterone, 0.625 mg conjugated estrogens daily and 200 mg micronized progesterone days 1–12.
5. Continuous combination estrogen and progestin, 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate daily.

The statistical analysis indicated that an increase in HDL-cholesterol levels was greater with unopposed estrogen (mean increase 5.6 mg/dL) and with sequential estrogen and micronized progesterone (mean increase 4.1 mg/dL), than in women receiving estrogen and medroxyprogesterone. However, as in the trials reviewed above, the decrease in LDL-cholesterol was essentially the same in all treatment groups, and significantly improved compared with the placebo group. There were no significant differences in systolic blood pressure, although all groups gradually increased systolic blood pressure with aging. There were no differences in diastolic blood pressure, and no changes with aging. The placebo group demonstrated the well-recognized increase in fibrinogen (a risk factor for cardiovascular disease) that occurs after menopause. All treatments prevented this increase in fibrinogen.

All treatment groups exerted a modestly favorable effect on fasting insulin and glucose levels, and the effect was greater in individuals with elevated fasting insulin concentrations. 428 These changes are consistent with a beneficial impact on insulin resistance, and most notably, the addition of a progestational agent, either medroxyprogesterone acetate or micronized progesterone, did not alter the responses.

The PEPI trial demonstrated a favorable impact on cardiovascular risk factors in women taking estrogen as well as combinations of estrogen and progestins. A major question is whether it is justified to interpret the results as indicating that a combination of estrogen plus micronized progesterone is clinically better because of the HDL-cholesterol responses. Studies have indicated that in women there is a 2–3% decrease in heart disease with a 1 mg increase in HDL-cholesterol. If this is accurate, then potentially the difference between estrogen and micronized progesterone compared with estrogen plus medroxyprogesterone acetate (a 3 mg HDL-cholesterol difference) would amount to a 9% difference in the risk of heart disease. However, the story is not that simple. Many of the beneficial actions of estrogen in terms of cardiovascular disease are dynamic effects on blood flow, atherosclerosis, the clotting system, and insulin levels. Given this complicated and multifactorial impact, the small difference in HDL-cholesterol is reduced in importance. It is worth emphasizing that the PEPI trial revealed the same beneficial effects in all treatment groups on HDL-cholesterol, fibrinogen, insulin levels, and glucose levels. In a Finnish ultrasonographic study, women receiving estradiol valerate and levonorgestrel had as great a reduction in large artery atherosclerotic plaques as those women on unopposed estrogen, despite a similar attenuation in HDL-cholesterol levels as observed in the PEPI trial. 429

One of the important contributions of PEPI is to provide a good response to the criticism that the estrogen users have less cardiovascular disease because clinicians give estrogen to healthier women. The results of this randomized trial demonstrate that estrogen has a greater impact compared with a placebo group, thus the design compensated for this criticism.

Clinical Events — Epidemiologic Data

The various mechanisms by which estrogen protects the cardiovascular system can be viewed as "surrogate end points," parameters by which the risk of cardiovascular disease can be assessed. A summary of the responses of these surrogate end points to the addition of progestational agents in estrogen regimens indicates a mixed story. This less than clear picture, we would argue, makes it impossible to determine with surrogate end points the overall effect of progestins on actual clinical events; e.g., myocardial infarctions. We must turn to epidemiologic data regarding actual clinical events, and here the data (just now emerging because estrogen and progesterin use is relatively recent) are reassuring.

Progestin Effects on Estrogen Actions

ATTENUATION

1. Increase in HDL and triglyceride levels.
2. Measure of acute reactivity; e.g., acute vasodilatation.
3. Cardiac inotropic activity; compliance responses.
4. Effect of estrogen on coronary atherosclerosis in the monkey.

NO ATTENUATION

1. Decrease in cholesterol and LDL.
2. Increase in fibrinolysis.
3. Inhibition of atherosclerosis (intimal thickening and remodeling).
4. Antioxidant activity.
5. Reduction in angiotensin converting enzyme (ACE) activity.

UNCERTAIN

1. Reduction in hyperinsulinemia.
2. Favorable changes in P-seletin.
3. Protection of endothelial cells.
4. Inhibition of macrophage foam cell formation.

A report from the ongoing cohort study from Uppsala, Sweden, provided information from the follow-up of approximately 23,000 women who were prescribed hormone therapy. 423 Overall, there was a 30% reduction in myocardial infarction in women prescribed estradiol valerate or conjugated estrogens. What is especially noteworthy was a 25% reduced risk of myocardial infarction in women exposed to a sequential estrogen-progestin regimen consisting of 2 mg estradiol valerate and 10 days each month of levonorgestrel (250 µg). Treatment with this combined estrogen-progestin product was also associated with a 40% reduced risk of stroke. 424

In an updated assessment of the Uppsala cohort, reduced mortality rates were documented for ischemic heart disease and intracerebral hemorrhage. 425 Once again, an even greater reduction in the risk of mortality from these conditions was observed in those women taking an estrogen-progestin combination.

In a case-control study from Seattle, unopposed estrogen and sequential estrogen and progesterin were associated with a 28% reduction in the risk of myocardial infarction, and the reduction increased with increasing duration of treatment in the current users. 426 There certainly was no evidence of attenuation secondary to the addition of a progestin. A retrospective cohort study from this same Seattle group determined prognosis in women surviving a myocardial infarction and detected a 36% reduced risk of re-infarction in current estrogen users, with a similar result in the one-third of their women on hormone therapy taking both estrogen and a progestin. 427

In a comparison of postmenopausal women in Finland receiving unopposed estrogen or a sequential regimen of estradiol valerate and levonorgestrel (250 µg for 10
days), both treatment groups had an equal reduction in ultrasonographically observed atherosclerotic plaques in the aorta and in the carotid and iliac arteries.

Carotid arterial wall thickening and the prevalence of carotid stenosis can be assessed by ultrasonography. In a study unique because it consisted of women who were 65-years old and older, the same beneficial decreases in carotid atherosclerosis could be documented in users of estrogen alone compared with women who were current users of estrogen and progesterin. In this study of elderly long-term users of postmenopausal hormone therapy, the HDL-cholesterol levels were similar in the two groups of women. This raises the possibility that the attenuation of estrogen-induced increases in HDL-cholesterol attributed to progesterin treatment may disappear with very long-term use of combined therapy.

As epidemiologic studies gather growing numbers of women on estrogen and progesterin, there is consistent reporting of an even greater beneficial impact with combined estrogen and progesterin than with estrogen alone. In the Nurses’ Health Study report on mortality, the relative risk of death for current users of estrogen-progesterin was 54% reduced compared with a 31% reduction with unopposed estrogen. In the 16-year follow-up report of the Nurses’ Health Study, the risk of coronary heart disease was reduced 40% in estrogen users and 61% in users of estrogen and progesterin. An explanation for this difference is not readily apparent.

The Heart and Estrogen/Progestin Replacement Study (HERS)

The HERS trial was a randomized, double-blind, placebo-controlled clinical trial 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. 2763 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. However, over time, differences were recorded.

<table>
<thead>
<tr>
<th>Year</th>
<th>Treated</th>
<th>Placebo</th>
<th>Relative Hazard (Risk) &amp; Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>57 cases</td>
<td>38 cases</td>
<td>1.52 (1.01–2.29)</td>
</tr>
<tr>
<td>Year 2</td>
<td>47 cases</td>
<td>48 cases</td>
<td>1.00 (0.67–1.49)</td>
</tr>
<tr>
<td>Year 3</td>
<td>35 cases</td>
<td>41 cases</td>
<td>0.87 (0.55–1.37)</td>
</tr>
<tr>
<td>Year 4</td>
<td>33 cases</td>
<td>49 cases</td>
<td>0.67 (0.43–1.04)</td>
</tr>
</tbody>
</table>

Thus, there was an increase in events in the first year (mostly in the first 4 months), and after two years of treatment, the appearance of a beneficial impact (although the annual relative risks did not achieve statistical significance, the test for the trend was significant). The authors attribute the increasing impact noted with increasing duration of treatment to a favorable effect on lipids, a 11% decrease in LDL-cholesterol and an 10% increase in HDL-cholesterol after one year, compared with the levels in the placebo group.

The results of the HERS trial were surprising in view of the overwhelming evidence from observational studies that postmenopausal hormone therapy prevents coronary heart disease. The authors offered two possible explanations for the difference between the HERS trial and previous observational studies. First, they pointed out the common and favorite criticism of the observational studies: selection bias, specifically that healthier women choose to use postmenopausal hormone therapy and therefore develop less coronary heart disease. This issue was addressed earlier in this chapter, noting that several studies have adjusted for this factor and a favorable impact of hormone therapy on coronary heart disease was still present. Thus, the common and favorite criticism of the observational studies: selection bias, specifically that healthier women choose to use postmenopausal hormone therapy and therefore develop less coronary heart disease. This issue was addressed earlier in this chapter, noting that several studies have adjusted for this factor and a favorable impact of hormone therapy on coronary heart disease was still present.

The authors further emphasized that the HERS trial consisted of older women with significant coronary heart disease, whereas observational studies have focused on primary prevention in younger and healthier women. However, there are observational studies that examined the impact of hormone therapy in women with pre-existing disease, and reported a beneficial effect. In the Leisure World study, estrogen users with previous myocardial infarctions, strokes, or hypertension had a 50% reduction in risk for death from a subsequent stroke or myocardial infarction. In the Lipids Research Clinics study, the cardiovascular mortality in women with previous cardiovascular disease was reduced 85%. And most impressively, in women with severe coronary disease (documented by arteriography), estrogen users had a 97% survival rate at 5 years compared with a significantly different 81% rate in nonusers. Estrogen therapy reduces the rate of restenosis in women who have undergone either coronary angioplasty or percutaneous atherectomy. In women who have undergone coronary artery bypass surgery, the 10-year survival rate in estrogen users was 81.4% compared with 65.1% in nonusers. In women who have been treated with estrogen after coronary angioplasty, case-control analysis indicated that the treated women had a better survival rate and experienced fewer subsequent myocardial infarctions.

The HERS trial is also in disagreement with recent reports that estrogen treatment could produce a regression in atherosclerosis. Imaging studies have documented a reduction in intimal thickening in postmenopausal women who are estrogen users compared with nonusers, and this beneficial effect was not compromised by the addition of a progestational agent to the treatment regimen. Thus, postmenopausal hormonal therapy has been reported to bring about a reduction in atherosclerosis, and this effect is comparable to that produced by a lipid-lowering drug.

The size of the HERS trial was determined by power calculations based upon specific assumptions. A comparison of the pre-trial assumptions with the actual outcomes reveals important differences:

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical event rate in the placebo group</td>
<td>5% /year</td>
</tr>
<tr>
<td>Drop-out rate</td>
<td>5% in year 1</td>
</tr>
<tr>
<td>Conversion of placebo to treatment rate</td>
<td>1% /year</td>
</tr>
<tr>
<td>Average follow-up</td>
<td>4.75 years</td>
</tr>
<tr>
<td>Recruitment</td>
<td>Paced</td>
</tr>
</tbody>
</table>

These differences between the assumptions and the actual events raise the important question whether the trial achieved sufficient statistical power to provide confident conclusions. The authors indicated that this concern was “partially” compensated by an 18% increase in recruitment. But, here too, there was a problem.

The reason why the average follow-up was less than expected is that most of the women were enrolled toward the end of the recruitment period. This seriously affects the accuracy of the impact of duration of treatment, an apparent beneficial effect in the HERS trial. Indeed, the late recruitment into HERS would have compensated for the lower average follow-up.

The Heart and Estrogen/Progestin Replacement Study (HERS) was a randomized, double-blind, placebo-controlled clinical trial 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. Overall, there were 172 myocardial infarctions and coronary deaths in the hormone group and 176 in the placebo group, obviously no difference. However, over time, differences were recorded.

<table>
<thead>
<tr>
<th>Year</th>
<th>Treated</th>
<th>Placebo</th>
<th>Relative Hazard (Risk) &amp; Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>57 cases</td>
<td>38 cases</td>
<td>1.52 (1.01–2.29)</td>
</tr>
<tr>
<td>Year 2</td>
<td>47 cases</td>
<td>48 cases</td>
<td>1.00 (0.67–1.49)</td>
</tr>
<tr>
<td>Year 3</td>
<td>35 cases</td>
<td>41 cases</td>
<td>0.87 (0.55–1.37)</td>
</tr>
<tr>
<td>Year 4</td>
<td>33 cases</td>
<td>49 cases</td>
<td>0.67 (0.43–1.04)</td>
</tr>
</tbody>
</table>

Thus, there was an increase in events in the first year (mostly in the first 4 months), and after two years of treatment, the appearance of a beneficial impact (although the annual relative risks did not achieve statistical significance, the test for the trend was significant). The authors attribute the increasing impact noted with increasing duration of treatment to a favorable effect on lipids, a 11% decrease in LDL-cholesterol and an 10% increase in HDL-cholesterol after one year, compared with the levels in the placebo group.

The reason why the average follow-up was less than expected is that most of the women were enrolled toward the end of the recruitment period. This seriously affects the accuracy of the impact of duration of treatment, an apparent beneficial effect in the HERS trial. Indeed, the late recruitment into HERS would have compensated for the lower average follow-up.

The increase in events recorded in the first year is difficult to understand. The most attractive explanation is to attribute, as the authors did, the increase to prothrombotic effects of estrogen. However, there is a host of evidence that indicates that postmenopausal hormone therapy (with and without progesterin) affects clotting factors in a pattern that favors fibrinolysis, an effect that should protect against thrombosis (reviewed earlier in this chapter). It may be that older women with atherosclerosis respond differently than younger, healthier women. In other words, increasing age and disease may change a woman’s thrombotic sensitivity to estrogen, perhaps because of the presence of unstable atherosclerotic plaques.

The most disturbing thought about the HERS trial is that the results reflect the impact of the continuous presence of medroxyprogesterone acetate. The clotting factor studies fail to find a detrimental effect of medroxyprogesterone acetate. However, there is reason to believe that the continuous presence of medroxyprogesterone acetate could attenuate and even block the favorable effects of estrogen on atherosclerosis and vasomotor function. There is evidence in the monkey that medroxyprogesterone antagonizes the favorable impact of conjugated estrogens on both the process of atherosclerosis and vasodilatation, but progesterone did not interfere with the ability of estrogen to inhibit atherosclerosis. Could attenuation of these estrogen effects make women with coronary heart disease more sensitive to any thrombogenic or ischemic effect of estrogen? In a study of mechanisms involved in the regression of atherosclerosis, conjugated estrogens did exert...
favorable activity (aortic connective tissue remodeling in response to lipid lowering) in the monkey, but medroxyprogesterone acetate prevented this action. Do the HERS results indicate a difference between sequential and daily, continuous regimens of estrogen-progestin treatment? The observational data have consistently failed to find any evidence that the addition of a progestational agent to estrogen therapy produced an attenuating impact on protection against clinical coronary events. However, this evidence is virtually all derived from experience with sequential regimens because the use of the daily, continuous methods is too recent for epidemiologic study. In our view, the HERS results combined with the experimental data derived from monkeys do raise concern. There are other treatment options, and products with new combinations are around the corner. At this point in time, treatment decisions must be informed clinical judgements, one of the reasons patients turn to their clinicians for assistance, an enduring and rewarding feature of the practice of medicine. All of the available evidence must be included in this decision making, not just the results of the HERS trial, and this evidence must be balanced with individual patient needs.

**Protection Against Cardiovascular Disease Is the Major Benefit of Hormone Therapy**

The uniformity and consistency of the literature on the effect of postmenopausal hormone therapy on the risk of cardiovascular disease are very impressive. All the population-based case-control studies and (with the re-analysis of the Framingham Heart Study) all of the prospective studies conclude that postmenopausal use of estrogens protects against cardiovascular disease. Sophisticated assessment and analysis (using the methods of information synthesis and meta-analysis) indicate that the effect of estrogen on heart disease is not controversial or ambiguous, but there clearly exists a protective benefit. We have such uniformity and consistency among the epidemiologic studies that the argument is very convincing. There remains, therefore, the important question of the nature and degree of impact due to the addition of progestational agents.

Epidemiologic evidence is accumulating to indicate that a postmenopausal dose that protects the endometrium avoids significant attenuation of estrogen's cardiovascular benefits. These epidemiologic data, however, reflect largely sequential regimens, and soon, we anticipate the appearance of data derived from women using daily, continuous combined methods. Ultimately, the best information will be derived from large, randomized trials, specifically the ongoing Women's Health Initiative in the U.S. (to be completed in 2008) and the WISDOM trial (Women's International Study of Long-Duration Oestrogen use after Menopause) in the U.K. (to be completed in 2011).

In view of the impressive association between estrogen and cardiovascular health in women, it is not illogical to ask whether a little estrogen therapy would be good for men. The normal levels of estrogen in men are important. The experimental suppression of normal physiologic levels of estradiol in young men leads to a significant decrease in apoprotein A-I and HDL-cholesterol, particularly the HDL 

The acute administration of estradiol into the coronary arteries of men, however, did not produce the vasodilatation and increased flow in response to acetylcholine that have been observed in women studied in a similar fashion. The failure of this acute response reflects the presence of androgens; e.g., androgen depletion of estrogen receptors. Given the major impact into the cardiovascular arteries of men, it makes sense that pharmacologic long-term treatment with a small dose of estrogen might have an impressive beneficial impact. Perhaps the administration of phytoestrogens to men will be an acceptable preventive health care approach.

**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 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 osteoclastic and osteoblastic 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. Calcium increases and PTH levels return to normal. A deficiency in estrogen is associated with a greater responsiveness of bone to PTH. Thus, for any given level of PTH 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. Osteopenia is sometimes used to indicate low bone mass, whereas osteoporosis is reserved for low bone mass with fractures. Osteoporosis is a major global public health problem, and it is epidemic in the United States, presently affecting more than 20 million individuals. 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. Other contributing factors include a dietary decreased in calcium and an earlier and greater loss of bone because of the impact of smoking. The Stone Age predecessors consumed a diet high in calcium, mostly from vegetable sources. 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.

**Pathophysiology**

Osteoporosis is characterized by 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 will depend on bone mass at the time of menopause and the rate of bone loss following menopause. Although the peak bone mass is influenced by 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 will be accumulated in young women by late adolescence (age 18), and the years immediately following menarche (11–14) are especially important. After adolescence, there continues to be only a slight gain in total skeletal mass that ceases around age 30, and in many individuals a decline in bone mass in the hip and spine begins after age 18. After age 30 in most people, there is a slow decline in bone mass density of approximately 0.7% per year. The importance of a normal diet and a normal hormonal environment during adolescence cannot be overstated.

The onset of spinal bone loss begins in the 20s, but the overall change is small until menopause. 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 four to eight times as fast as cortical bone. Beyond age 30, trabecular resorption begins to exceed formation by about 0.7% per year. This adverse relationship accelerates after menopause and up to 5% of trabecular bone and 1–1.5% of total bone mass will occur per year after menopause. This accelerated loss will continue for 10–15 years, after which bone loss is considerably diminished but continues as the natural loss. For the first 20 years following cessation of menses, menopause-related bone loss results in a 50% reduction in trabecular bone and a 30% reduction in cortical bone.

When estrogen levels decline, bone remodeling increases. Each remodeling unit is initiated by osteoclast excavation followed by osteoblast resurfacing. 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. Increased efficiency of calcium absorption (probably secondary to estrogen-induced enhancement of the availability of vitamin D (1,25-dihydroxycholecalciferol) 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. Estrogen modulates the production of bone resoring cytokines such as interleukin-1 and -6, bone stimulating factors such as insulin-like growth factors I and II, and transforming growth factor-b. Estrogen increases vitamin D receptors in...
osteoblasts, and this may be a method by which estrogen modulates vitamin D activity in bone.440 There is little evidence that estrogen affects bone by altering the circulating calcitropic hormones.441 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.442, 443 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.444 And most impressively, a man with an aromatase deficiency, treated with estrogen, demonstrated that both androgens and estrogens are necessary in order to for males to reach optimal bone mass.445

The process is slower in blacks, and 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.446, 447 The reasons for this are not entirely clear; it is not simply a difference in estrogen metabolism.448

In general, bone mass is increased in black and obese women and decreased in white, Asian, thin, and sedentary women. Despite the greater bone mass (and lower fracture rates) in blacks, the impact of osteoporotic fractures in black women is still considerable. Black women are subject to similar risk factors for osteoporosis (thinness, high alcohol consumption), and, most importantly,

postmenopausal estrogen therapy is associated with protection.449 Japanese women demonstrate the same amount and pattern of bone loss after menopause as white women.450

Although estrogen plays a principal role in regulating bone density, a genetic susceptibility 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.451 Studies of twins and mother-daughter pairs indicate that up to 70% of the variation in bone density is determined by heredity.452

Variations in the gene that encodes the vitamin D receptor are prevalent in postmenopausal women with decreased bone densities.453 The absence of vitamin D receptor gene allelic polymorphisms has been reported in older women who neither lose significant bone nor respond to calcium supplementation.454 In addition, low bone density is associated with specific alleles of COLIA1, one of the two genes that encode the two polypeptides of collagen.455 This opens the way to develop genetic markers to identify individuals at increased risk for osteoporosis, and, ultimately, to develop genetic therapy.

The change in trabecular 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 attributable to estrogen deficiency rather than to aging itself.456, 457 Vertebral bone is especially vulnerable, beginning to decline as early as 20 years of age.458 Vertebral bone mass can be found to be significantly decreased in perimenopausal and early postmenopausal women who have rising FSH and decreasing estrogen levels, while bone loss from the radius is not found until at least a year out from the menopause.459 This early loss of axial skeleton bone suggests that the hypoestrogenic postmenopausal state is not the only cause of vertebral osteoporosis.460 One obvious suspect is a decline in dietary intake of calcium 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 upon 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.461

Bone Loss During the Perimenopausal Transition

Should a clinician be concerned about bone loss and consider intervention during the perimenopausal years? Some studies have concluded that calcium supplementation of perimenopausal women retards metacarpal and lumbar bone loss.462, 463 However, the amount of perimenopausal bone loss is small unless estrogen levels are below normal.464, 465 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.466, 467 Interventions and treatments to prevent future osteoporosis are not necessary in women who have adequate estrogen levels and who are eating normally.

Signs and Symptoms

The osteoporotic disabilities sustained by the castrate or postmenopausal woman include back pain, decreased height and mobility, and fractures of the vertebral
body, humerus, upper femur, distal forearm, and ribs. Back pain is a major clinical symptom of vertebral compression fractures. The pain with a fracture is acute, and then it subsides over 2–3 months, but lingers as chronic low back pain due to increasing lumbar lordosis. The pain will subside within 6 months unless multiple fractures produce a picture of constant pain.

Epidemiologic studies have revealed the following:

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 5 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 non-treated 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 3 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 0.3/1000 to 20/1000 from 45 to 85 years. Eighty percent of all hip fractures are associated with osteoporosis. White women have approximately a 16% lifetime risk of having a hip fracture. This fracture carries an increased risk of morbidity and mortality. Between 15–20% of patients with hip fracture die due to the fracture or its complications (surgical, embolic, cardiopulmonary) within 3 months, and Beals found that only half of hip fracture victims in Portland, Oregon, survived for 1 year. However, most of this increase in mortality is due to underlying conditions and prevention of these fractures would have only a minor impact on longevity. Nevertheless, the survivors are frequently severely disabled and may become permanent invalids. Hip fractures alone occur in more than 300,000 women per year in the U.S. with a mortality of 40,000 annually and an associated cost of billions of dollars.

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. Even in women without osteoporosis, there is a correlation between spinal bone density and number of teeth.

### 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.

### Summary of Reasons to Measure Bone Mass

1. To help patients to 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-aging treated long-term with glucocorticoids, thyroid hormones, 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.

### Specific Causes of Bone Disease

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Endocrine diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heparin, anticoagulants, high intake of alcohol</td>
<td>Excess glucocorticoids</td>
</tr>
<tr>
<td>Renal and hepatic</td>
<td>Hyperparathyroidism</td>
</tr>
<tr>
<td>Calcium, phosphorous, vitamin D deficiencies</td>
<td>Estrogen deficiency</td>
</tr>
<tr>
<td></td>
<td>Hyperparathyroidism</td>
</tr>
</tbody>
</table>

### 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). Measurement of lower bone mass in the hip is even more predictive; a one standard deviation is associated with nearly a 3-fold increase in risk of fracture. 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 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 and single-energy x-ray absorptiometry because measurement of bone loss at the heel and radius accurately assess future fracture risk.

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. 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, thyroidone, anticonvulsant, or heparin treatment deserve bone mass assessment.

There is a percentage of postmenopausal women on hormone therapy (from 5% to 15%, depending upon compliance) who continue to lose bone. Consideration should be given to an occasional measurement of bone density as an effective method of assessment and to motivate compliance.
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. A dual-energy absorptiometry employs 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 for bone density measurements can be performed on most commercial computed tomography (CT) 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 three sites of greatest interest, the radius, the hip, and the spine. * Better accuracy is gained by 3-site assessment 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. Serial measurements are usually at least one year apart. For practical clinical use (and for screening), measurement of the bone density at the radius or calcaneus is sufficient and cost-effective. It is anticipated that ultrasonography will prove to be a low-cost effective method for bone mass assessment. 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. Ultrasonometry of the calcaneus can also predict spine and femur fracture risk.

### T-Z Score Definitions

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to -1 S.D. from the reference standard (84% of the population)</td>
<td>Standard deviations between patient and average peak young adult bone mass. The more negative, the greater the risk of fracture.</td>
</tr>
<tr>
<td>-1 to -2.5 S.D.</td>
<td>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.</td>
</tr>
</tbody>
</table>

### 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. Because of individual variability, marker measurements correlate poorly with bone mineral density; however, changes in these markers quickly indicate responses to estrogen therapy. Therefore, efficacy of therapy (any of the anti-resorptive treatments) can be judged by comparing a baseline value with a single follow-up measurement after 1–3 months of treatment. 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 yet to be determined.

### Treatment

Estrogen therapy will stabilize the process of osteoporosis or prevent 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. When estrogen is supplemented with calcium, an 80% reduction in vertebral bone loss can be observed in vertebral compression fractures can be observed. This reduction is seen primarily in patients who have taken estrogen for more than 5 years.

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 7th and 8th 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. 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. 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. In a Swedish case-control study, most of the beneficial effect of hormone therapy was lost 5 years after discontinuing treatment. Maximal protection against osteoporotic fractures, therefore, requires lifelong therapy, and even some long-term protection requires 10 or more years of treatment. Standard doses of estrogen administered transdermally (50 µg) appear to protect against fractures as well as standard oral doses.

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 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. 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. It is still unknown whether a continuing increase over many years provides additional fracture protection.

The positive impact of hormone therapy on bone has been demonstrated to take place even in women over age 65. 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. Serial measurements are usually at least one year apart. For practical clinical use (and for screening), measurement of the bone density at the radius or calcaneus is sufficient and cost-effective. It is anticipated that ultrasonography will prove to be a low-cost effective method for bone mass assessment. 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. Ultrasonometry of the calcaneus can also predict spine and femur fracture risk.

How reliable, practical, and cost-effective this approach is for routine clinical use is yet to be determined.

The positive impact of hormone therapy on bone has been demonstrated to take place even in women over age 65. 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. Serial measurements are usually at least one year apart. For practical clinical use (and for screening), measurement of the bone density at the radius or calcaneus is sufficient and cost-effective. It is anticipated that ultrasonography will prove to be a low-cost effective method for bone mass assessment. 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. Ultrasonometry of the calcaneus can also predict spine and femur fracture risk.

How reliable, practical, and cost-effective this approach is for routine clinical use is yet to be determined.
The reason why some women fail to respond is unknown. It is
likely that the addition of medroxyprogesterone acetate or micronized progesterone. Thus, the daily, continuous combination of estrogen-progesterin is equally efficacious in maintaining bone density as the standard sequential regimens, although at least one study has indicated a greater response in the lumbar spine with continuous treatment.

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. 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.

Follow-up Assessment of Hormone-Treated Women

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. 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.

It is worthwhile to measure the bone density in treated women when they are in their late 60s. The reason why some women fail to respond is unknown. It is further unknown whether these patients will respond to added treatment, such as calcitonin or a bisphosphonate, but it is worth special evaluation, treatment, and surveillance.

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. Raloxifene exerts no proliferative effect on the endometrium but produces favorable responses in bone and lipids. The changes in bone remodeling produced by raloxifene are consistent with an estrogen agonist effect.

Estrogen selective agonists/antagonists (tamoxifen and raloxifene) have been reported to reduce the incidence of breast cancer within 2–3 years of treatment. Short-term clinical trial data indicate that raloxifene has a positive impact on bone density, favorable effects on LDL-cholesterol, fibrinogen, and lipoprotein(a), but no effects on HDL-cholesterol and PAI-1, and raloxifene increases hot flushing. Long-term clinical trial data will be necessary to determine the ultimate impact on clinical events, specifically fractures, coronary heart disease, stress incontinence, and cognition. The MORE study (Multiple Outcomes of Raloxifene Evaluation) is a worldwide clinical trial of raloxifene in osteoporotic women. After 2 years, raloxifene treatment was associated with a 44% reduction in vertebral fractures. 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. However, a combination of actions (antioxidant activity, some beneficial effects on lipids, a reduction in homocysteine levels) makes it likely there will be some favorable impact on the cardiovascular system. Raloxifene Use for The Heart is both a primary and secondary coronary heart disease prevention trial, started in 1998, that should provide data by 2003. In our view, raloxifene is an option for prevention of osteoporosis, especially for patients reluctant to use hormone therapy, but not a substitute for estrogen.

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 have been 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 involve women beyond this early stage of the postmenopausal period definitely indicate a positive impact of calcium supplementation.

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 (1000 mg per day) reduces bone loss and decreases fractures, especially in individuals with low daily intakes. However, 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. In order to remain in zero calcium balance, women on estrogen therapy require a total of 1000 mg elemental calcium per day. Because the average woman receives about 500 mg of calcium in her diet, the minimal daily supplement equals an additional 500 mg. Women not on estrogen require a daily supplement of at least 1000 mg calcium.

Even with the commonly used therapeutic doses of calcium, nearly 40% of postmenopausal women will have inefficient absorption. 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. Nevertheless, the beneficial impact of estrogen on bone is reduced in the absence of calcium supplementation.

Improved calcium intake in adolescents results in significant increases in bone density and skeletal mass, providing protection against osteoporosis later in life. 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 1500 mg. This amount, 1500 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.

<table>
<thead>
<tr>
<th>Calcium Content of Foods</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yogurt (one cup)</td>
<td>415 mg</td>
<td></td>
</tr>
<tr>
<td>Yogurt with fruit (one cup)</td>
<td>345 mg</td>
<td></td>
</tr>
<tr>
<td>Juice fortified with calcium (one cup)</td>
<td>300 mg</td>
<td></td>
</tr>
<tr>
<td>Milk (one cup)</td>
<td>300 mg plus 100 IU Vitamin D</td>
<td></td>
</tr>
</tbody>
</table>
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, and calcium gluconate only 9%. The calcium carbonate antacids are excellent, inexpensive sources. One should 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 can be contaminated with lead. 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.

### Vitamin D

Osteoporosis related to aging is due significantly to age-related changes in vitamin D and calcium metabolism. There is an age-related decrease in the ability of the skin and the kidney to synthesize the active form of vitamin D (1,25-dihydroxycholecalcifer, also 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 cholecalcifer (vitamin D). Commercial vitamin preparations contain ergocalcifer (vitamin D or calcifer). Vitamin D-fortified milk contains cholecalcifer. Cholecalcifer and ergocalcifer are converted in the liver to calcifolder, which is converted in the kidneys to calcitriol (1,25-dihydroxyvitamin D).

It is now recommended that individuals over age 70 should add 800 units of vitamin D to calcium supplementation. A large randomized trial in Finland documented a reduced rate of fractures in elderly women receiving supplementation of vitamin D by (an annual intramuscular injection), and in France, supplementation of calcium and vitamin D reduced the number of hip fractures by 43%. In postmenopausal women with osteoporosis, treatment with calcitriol (vitamin D) produced a 3-fold reduction in vertebral fractures. And finally, supplementation with 400 IU vitamin D in women age 70 and older restored the markers of vitamin D and calcium metabolism to that observed in young adults, and the bone mineral density at the femoral neck increased.

Because adequate and active vitamin D depends upon cutaneous generation mediated by sun exposure, women who live in cloudy areas during the winter months are relatively vitamin D deficient and lose bone. In far northern and southern areas, the winter sunlight is inadequate to stimulate dermal activation. Vitamin D supplementation is recommended for these women as well but only during the winter and at a lower level, 400 units daily (usually available in over-the-counter multivitamins or in combination with calcium). High doses cause hypercalcemia with a risk of stone formation. If uncertain regarding vitamin D supplementation, the serum level of 1,25-dihydroxyvitamin D can be measured; a level below 15 ng/L is abnormal.

The impact of vitamin D supplementation may be less in younger, early postmenopausal women. In a 2.5-year randomized clinical trial in Finland, supplementation with 300 IU vitamin D did little by itself and added nothing to the beneficial effect of estrogen on bone density. The benefit of vitamin D supplementation is clear in older women, and the lack of side effects with low doses encourages us to recommend vitamin D supplementation as part of the overall program for osteoporosis prevention in younger women.

### Bisphosphonates

Bisphosphonates are effective in preventing bone loss by inhibiting bone resorption. The bisphosphonates bind to bone mineral where they remain for many years, making bone less susceptible to osteoclastic action. The most promising of these agents are etidronate, alendronate, clodronate, pamidronate, tiludronate, and risedronate. 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.

In women with osteoporosis, alendronate (10 mg daily) administration reduced the risk of subsequent non-vertebral fractures by at least 30% (and probably 50%) and vertebral fractures by 90% in the first 3 years of treatment. 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. The increase in bone density with the 5 mg dose is slightly less than that observed with estrogen-progesterin therapy. It is unlikely that a difference of a few percentage points in bone density gain has an impact on the number of fractures ultimately experienced. Alendronate must be taken on an empty stomach with a full glass of water at least 30 minutes before any other food or liquid intake in order to achieve adequate absorption. A failure to remain upright for at least 30 minutes and until after the first food intake of the day after ingesting alendronate can result in esophageal injuries, such as esophagitis, esophageal ulcers, and esophageal erosions with bleeding. Risedronate, 5 mg daily, is as effective as alendronate for the prevention of bone loss and may be better tolerated.

Etidronate given in a cyclic fashion (e.g., 400 mg daily for 2 weeks followed by a 12-week drug-free interval during which calcium supplementation is administered) prevents bone loss and reduces the vertebral fracture rate. In many countries, etidronate is the drug of choice for the treatment of osteoporosis. Although no comparison studies are available, it is probable that the bisphosphonates are comparable in efficacy and safety.

Bisphosphonates are an effective addition to osteoporotic prevention because they are well tolerated if taken correctly. However, unlike estrogens, bisphosphonates are unidimensional, having no effect on cardiovascular disease, hot flushes, or the atrophic changes seen in menopause. A combination of bisphosphonates and estrogen has an additive effect on bone; with alendronate and hormone therapy, there was about a 2% greater gain in bone density in the spine after 2 years of treatment. It is not believed that this addition of several percentage points in bone mineral density will provide a worthwhile gain in fracture protection. Unlike, estrogen, accelerated bone loss does not occur after discontinuation of alendronate therapy.

Calcitonin

Calcitonin regulates plasma calcium by inhibiting bone resorption and can be used in patients for whom hormone therapy is contraindicated. Given by subcutaneous injection in a dose of 100 IU daily to women early after menopause it has the same effectiveness as estrogen in conserving bone density. Studies with intranasal delivery of salmon calcitonin (200 IU dily) indicate it can be similarly effective. Calcitonin treatment should be combined with vitamin D and calcium supplementation. The disadvantages are the high cost and the potential for immunologic reactions to non-human calcitonin. Human calcitonin is available, but recombinant salmon calcitonin is more potent. An oral preparation will soon be available.

### Fluoride

The addition of fluoride, a potent stimulator of bone formation, offers significant protection against osteoporosis. The clinical response depends on the formulation and dose. Slow-release sodium fluoride (25 mg bid given 12 of every 14 months) combined with calcium supplementation reduces the vertebral fracture rate with essentially no side effects. Treatments recommended for no longer than 4 years to avoid the toxic accumulation of fluoride in bone. Thus, this treatment is reserved for patients with established postmenopausal osteoporosis.
Tibolone

Tibolone is a steroid, related to the 19-nortestosterone family, that is effective for the treatment of bone and hot flushes in a dose of 2.5 mg per day. A lower dose of 1.25 mg daily also provides bone protection, but it is less effective and there is more vaginal bleeding. Tibolone is metabolized into 3 steroidal isomers with varying estrogenic, progestogenic, and androgenic properties. The metabolites differ in their activities and dominance according to the target tissue. Thus, tibolone provides estrogenic effects on bone and hot flushing, but it induces atrophy of the endometrium. Its beneficial impact on bone (2.5 mg dose) is comparable to standard hormonal therapy. In the endometrium, tibolone is converted locally (by endometrial 3 beta-hydroxysteroid dehydrogenase/isomerase) to its D4 progestational isomer, explaining its nonproliferative effect on the endometrium. Tibolone has an estrogenic effect on the vagina, and women report improvements in the symptoms of vaginal dryness and dyspareunia, and an increase in sexual enjoyment and libido.

Although a short-term reduction in HDL-cholesterol is an undesirable consequence; the long-term impact has not been well studied. In a 2-year study, the unfavorable effect on lipoproteins was accompanied by beneficial changes in coagulation factors consistent with enhanced fibrinolysis and unchanged coagulation. Overall, it is possible that some favorable activity on the cardiovascular system is maintained. A major advantage of tibolone (2.5 mg daily) is its low (10–20%) incidence of bleeding. Because tibolone inhibits breast cell proliferation in vitro, it is possible that future studies will indicate that tibolone offers some protection against breast cancer. Tibolone also has a beneficial impact in short-term studies on insulin resistance in normal women and in women with noninsulin-dependent diabetes mellitus.

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.

Growth Hormone

Human growth hormone administration stimulates bone remodeling, but does not produce a significant increase in bone density. The extreme expense and the high incidence of unpleasant side effects further make this option unattractive.

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.

Lifestyle Modifications

Lifestyle can have a beneficial effect on bone density. Physical activity (weight-bearing), as little as 30 minutes a day for 3 days a week, will increase the mineral content of bone in older women. To be effective, exercise must exert a load on bone, especially the spine. Ordinary walking will not suffice. Even brisk walking achieves a significant increase in bone density only in the calcaneus, the site subjected to stress with walking. However, brisk walking may slow the rate of bone loss in the hip. 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. 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 will improve balance and decrease the risk of falling. For these reasons, walking, even after adjusting for bone density, is associated with a reduced risk of hip fracture.

The impact of exercise on bone is significantly less than that achieved by hormone therapy. Women require the full combination of pharmacologic therapy, calcium 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.

Adverse habits, such as cigarette smoking or 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. Women who smoke also enter menopause earlier, and lose bone at a greater rate in the first years of the postmenopausal period. Studies have indicated that estrogen is associated with lesser protection against fractures in smokers; however, this may be correctable by increasing the blood level of estrogen with the dose administered. The lower blood levels of estrogen in smokers have been correlated with an earlier menopause and a reduced bone density, and, therefore, the standard dose of estrogen may not totally counteract the predisposition of smoking toward osteoporosis. 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 or urinary markers would further aid in achieving maximal effects of therapy.

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 estrogen/progestin therapy, and excessive thyroid effects can be avoided by annually monitoring treatment dosage with TSH levels. Specific treatment should also be offered to patients using anticonvulsants.

A high coffee intake has been reported to be associated with an increased risk of osteoporosis. However, this increase in risk is dependent upon 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. 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 increases bone density without an increase in weight or body fat.

Remember that not all fractures are solely due to osteoporosis. Drug side effects, impaired vision, neurological dysfunction, and muscular conditions all put patients at risk because more than 90% of fractures occur following a fall. 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. In addition, there is evidence that estrogen with or without added progestin improves muscle strength and balance. On the other hand, some studies have not been able to document an increase in muscle strength or improvements in balance. 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.

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


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. However, in the last decade, the focus of postmenopausal hormone therapy has 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. It is increasingly accepted that appropriate doses of estrogen have a beneficial impact on the risk of cardiovascular disease. The long-term imbalances on urinary incontinence and cognition remain to be documented, but there is reason to believe there will be benefits in these areas.

We suggest treatment with estrogen for all women disturbed by the symptoms of hormone deprivation and advocate hormonal prophylaxis against osteoporosis and cardiovascular disease. The decision to use or not to use estrogen belongs to the patient, and it should be based on the information available in this book. The recommendation that hormone therapy be given for the shortest period of time appears to be shortsighted in view of the impressive evidence that sustained therapy has a profound impact on osteoporosis and cardiovascular disease, and that there are more beneficial than potentially harmful effects.

The evidence supporting many of the benefits with postmenopausal hormone therapy is reviewed in the previous chapter (Chapter 17, Menopause and the Perimenopausal Transition) where the effects of hormonal treatment are considered in conjunction with the impact of postmenopausal estrogen deprivation. In this chapter, we will review the clinical aspects of postmenopausal hormone therapy and our methods of patient management.

History

Charles Edouard Brown-Sequard, practicing in London, reported in 1889, that he was rejuvenated by the self-administration of testicular extracts, and he suggested that ovarian extracts would have the same effect in women. Efforts to do so around the turn of the century were largely unsuccessful, but in 1897, ovarian extract was reported to be effective for hot flushing. 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.

In the 1930s, the ovarian hormones were isolated, and the “estrin” products and synthetic estrogens, stilbestrol and ethinyl estradiol, were administered to menopausal women. Allen and Doisy were the first to isolate the estrogenic ovarian hormone. Edgar Allen was born in Colorado, educated at Brown University, and served in France during World War I. In 1933, Allen became the chairman of the Department of Anatomy at Yale University. He died of a heart attack, on patrol off Long Island, for the U.S. Coast Guard, in February 1943.

Edward 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 Physiology or Medicine in 1955 for the isolation of the steroid hormones.
Nobel Prize in Medicine, along with Henrik Dam, in 1943, for his isolation and synthesis of vitamin K. He died in 1986 at the age of 92.

In 1919, Edgar Allen and Edward Doisy, being 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 driving to work together. In 1922, Allen moved to the University of Missouri to be Professor of Anatomy (and Dean in 1930), and Doisy 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, A.S. Parkes and C.W. Bellerby coined the basic word estrin to designate the hormone (or hormones) that induces estrus in animals. The terminology was extended to include estrone, estradiol, and estriol, although significant amounts of pure estradiol were not isolated until 1936. Doisy and his students (Veler and Thayer) in St. Louis isolated estrone in crystalline form from the urine of pregnant women in 1929. In 1932, at the first meeting of the International Conference on the Standardization of Sex Hormones in London, the pioneering chemists were bemoaning the problem of scarcity, limiting supplies to milligram amounts, when a relatively unknown biochemist, Girard, offered 20 grams of crystalline substance, derived by the use of a new reagent to treat mare's urine.

George W. Corner, at the University of Rochester, invited Willard Myron Allen, an organic chemist but now a medical student at the time in the 1920s, 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 a crystalline progestin 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.

Corner was born in Baltimore, graduated from the Johns Hopkins medical school, and completed an internship in gynecology under Howard Kelley. He joined the Department of Anatomy at the University of California, and in 1923 at the age of 34, he became the first professor of anatomy in the new medical school of the University of Rochester. In 1940, Corner became the Director of the Carnegie Laboratory of Embryology in Baltimore. He died in 1981.

Allen was born in Macedon, New York, educated at Hobart College and Brown University, finally receiving his medical degree from the University of Rochester in 1932. He completed a residency in obstetrics and gynecology and served on the faculty in Rochester until he became chairman of obstetrics and gynecology at the Washington University School of Medicine in 1940. He died in 1993.

In the 1930s, the Ayerst Company was extracting estrogens from the urine of pregnant women. Limited by the problems of supply, low activity, plus 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. 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.

### Composition of Conjugated Estrogens (Premarin)

<table>
<thead>
<tr>
<th>Estrogen</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium estrone sulfate</td>
<td>49.3%</td>
</tr>
<tr>
<td>Sodium equilin sulfate</td>
<td>22.4%</td>
</tr>
<tr>
<td>Sodium 17α-dihydroequilin sulfate</td>
<td>13.8%</td>
</tr>
<tr>
<td>Sodium 17α-estradiol sulfate</td>
<td>4.5%</td>
</tr>
<tr>
<td>Sodium D8,9-dehydroestrone sulfate</td>
<td>3.5%</td>
</tr>
<tr>
<td>Sodium equilenin sulfate</td>
<td>2.2%</td>
</tr>
<tr>
<td>Sodium 17β-dihydroequilin sulfate</td>
<td>1.7%</td>
</tr>
<tr>
<td>Sodium 17α-dihydroequilenin sulfate</td>
<td>1.2%</td>
</tr>
<tr>
<td>Sodium 17β-estradiol sulfate</td>
<td>0.9%</td>
</tr>
<tr>
<td>Sodium 17β-dihydroequilenin sulfate</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

Estrogens and progestins used for postmenopausal hormone therapy are among the most commonly prescribed medications in the U.S. Currently, 46% of women who have experienced a natural menopause and 71% of women who have had bilateral oophorectomy report having used postmenopausal hormone therapy. The average duration of use in the U.S. as of 1992 was 6.6 years, but only 20% of users had maintained treatment for at least 5 years.

### Selection of Patients for Treatment

#### Women Under Age 40 (Castrates and Women With Gonadal Dysgenesis)

In these women, the duration of estrogen deprivation is prolonged, and in women after surgical menopause the loss of estrogen is acute. The cyclic use of estrogen is recommended for short-term reduction of vasomotor symptoms and for long-term prophylaxis against cardiovascular disease, osteoporosis, and target organ atrophy. In some young patients, the equivalent of 0.625 mg conjugated estrogens is insufficient to allow menstrual bleeding. Because women of this age ordinarily are exposed to estrogen levels that stimulate endometrial growth and withdrawal bleeding, and for psychological reasons, a higher dose should be considered, if necessary, to maintain withdrawal bleeding until the menopausal time of life. However, in our experience, the doses equivalent to 0.625 mg conjugated estrogens almost always suffice. A standard sequential program is utilized. In those patients castrated because of endometriosis, recurrence of endometriosis has very rarely been a problem with estrogen therapy, but because endometrial cancer has been reported to occur in remaining endometriosis exposed to unopposed estrogen, an estrogen and progestin combination is strongly recommended.

### The Perimenopausal Transition Years

After exclusion of other gynecologic causes, dysfunctional uterine bleeding is treated by progestin or oral contraceptive therapy and, if necessary, biopsy surveillance. Vasomotor reactions appearing in women despite the presence of menstrual bleeding should receive careful evaluation. Abnormal thyroid function should be ruled out. If a follicle-stimulating hormone (FSH) level is not greater than 20 IU/L, serious consideration should be directed toward psychosocial reasons for the flushing response.

The risk of fracture from osteoporosis will depend on bone mass at the time of menopause and the rate of bone loss following menopause. 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 will be accumulated in young women by late adolescence (age 18), and the years immediately following menarche (11–14) are especially important.

Meaningful perimenopausal bone loss that is secondary to a decrease in estrogen is limited to those women with fluctuating hormonal and menstrual function, irregular bursts of follicular function alternating with no ovarian response to gonadotropins, so that function alen are exposed to low estrogen over significant periods of time. The bone loss that begins in the 20s is due to a mechanism that is independent of hormones, and longitudinal studies have documented that trivial amounts of bone are lost prior to menopause in women with adequate estrogen levels.

### The Early Postmenopause

The long-term postmenopausal use of hormone therapy depends heavily on a woman's own informed assessment, a process that should occur at this point in life. An...
understanding of hormone therapy is an important component in any preventive health program directed toward the menopausal years. As a result of immediate responses in early climacteric symptoms, the patient enters the postmenopausal period of life more confident of herself emotionally, sexually, and physically. In our view, this establishes or cement good patient-clinician interchange and relations. The follow-up of the patient on effective estrogen-progestin therapy is more secure and certain. The practitioner offering estrogen-progestin treatment has a better and more reliable opportunity to act as primary clinician for these aging women. All monitoring of health systems will be improved as a result of this single involvement.

The Late Postmenopause

Atrophic conditions can be effectively treated with local or oral therapy in low maintenance doses. If there is no apparent basis for osteoporosis other than aging and ovarian failure, estrogen-progestin therapy and calcium plus vitamin D supplementation are advisable even for very old women. Further loss of bone can be halted and the risk of fractures reduced. In these older women, an assessment of progress can be obtained by measuring bone density. The impact of initiating hormone therapy in elderly women on the risk of cardiovascular disease has not been ascertained. However, it makes sense that elderly women would benefit from estrogen’s dynamic protective mechanisms with an impact on stress incontinence, and probably, on the risk of Alzheimer’s disease, as well.

Hormonal Agents and Routes of Administration

The dose of estrogen that is maximally effective in maintaining the axial and peripheral bone mass is equivalent to 0.625 mg conjugated estrogens. The relative potencies of commercially available estrogens become of great importance when prescribing estrogen, and the clinician should be familiar with the potencies in the Table.

<table>
<thead>
<tr>
<th>Relative Estrogen Potencies</th>
<th>FSH levels</th>
<th>Liver proteins</th>
<th>Bone density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjugated estrogens</td>
<td>1.0 mg</td>
<td>0.625 mg</td>
<td>0.625 mg</td>
</tr>
<tr>
<td>Micronized estradiol</td>
<td>1.0 mg</td>
<td>1.0 mg</td>
<td>1.0 mg</td>
</tr>
<tr>
<td>Estropipate (piperazine estrogen sulfate)</td>
<td>1.0 mg</td>
<td>1.25 mg</td>
<td>1.25 mg</td>
</tr>
<tr>
<td>Ethinyl estradiol</td>
<td>5.0 µg</td>
<td>2–10 µg</td>
<td>5.0 µg</td>
</tr>
<tr>
<td>Estradiol valerate</td>
<td>—</td>
<td>—</td>
<td>1.0 mg</td>
</tr>
<tr>
<td>Esterified estrogens</td>
<td>—</td>
<td>—</td>
<td>0.625 mg</td>
</tr>
<tr>
<td>Transdermal estradiol</td>
<td>—</td>
<td>—</td>
<td>50 µg</td>
</tr>
</tbody>
</table>

The 17α-ethinyl group of ethinyl estradiol (by resisting metabolism) appears to enhance hepatic effects, for no matter by which route it is administered, liver function is affected. 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.

At least two studies have been unable to demonstrate prevention of bone loss with the administration of 2 mg estradiol daily. 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. Estril 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. In pregnancy, where the concentration of estril is very great, it can be an important hormone, not just a metabolite. Thus, higher estril levels are not necessarily protective. Because estril protects the rat against breast tumors induced by various chemical carcinogens, it has been hypothesized that a higher estril level protects against the more potent effects of estrone and estradiol. Indeed, antagonism of estradiol occurs only at very narrow range of the ratio of estradiol to estril, a range rarely encountered either physiologically or pharmacologically. Below this range, estradiol is unimpaired, above this range estril itself exerts estrogenic activity.

Esterified estrogens are synthetically prepared from plant precursors and are composed mostly of sodium estrone sulfatate with a 6–15% component of sodium equin sulfatate. Estradiol valerate is rapidly hydrolyzed to estradiol, and, therefore, the pharmacology and effects are comparable at similar dosages.

Transdermal 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. The current generation of patches has the hormones dissolved and distributed throughout the adhesive matrix. In a study of women who previously discontinued patches because of skin irritation (contact dermatitis), skin reactions were less common with the newer matrix patches. In addition, the matrix patches are better tolerated in tropical environments. The patches are designated according to the amount of estrogen delivered per day: 50 µg and 100 µg.

The effect of steroids on lipids and lipoproteins is determined by the type of steroid, the dose, and the route of administration. An obstacle to the use of transdermal hormone therapy has been the scarcity of data indicating a beneficial impact on the lipoprotein profile. There has been concern that delivery of estrogen through the skin yields a blood level that might be too low to provide protection against cardiovascular disease, especially because after peak concentrations in the first day after application, there is a progressive decrease that can be relatively rapid. Furthermore, there is marked variation in levels among individuals and within the same individual.

The concentration of estrogen in the portal system after oral administration is 4–5 times higher than that in the periphery. Furthermore, the estradiol:estrone ratio differs in the portal system. Thus, the first-pass effect is either significant for the lipoprotein effects, or it is important only in the short term. For example, short-term studies (6 weeks) could document increased catabolism of LDL and increased production of apoprotein A-I with oral estrogen, but no effect with transdermal estrogen. And a 2-year study in Los Angeles with a transdermal dose (100 µg) detected no significant change in HDL-cholesterol levels. 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. Of note, 12% of the women on either the transdermal or oral regimen lost bone from the femoral neck, despite documented adequate compliance. The transdermal administration of 100 µg estradiol combined with a progestin will not only increase bone density, but also reduce the fracture rate in older women who already have significant osteoporosis. However, standard doses of estrogen administered transdermally (50 µg) appear to protect against fractures as well as standard oral doses.

Until data are available documenting the degree of impact of the various routes of administration on actual clinical events (especially cardiovascular disease), the prudent clinical decision is to select the method (an oral program) that has epidemiologic support.

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. 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. Progestational treatment is necessary, and because of the higher blood levels, a minimal duration of 14 days each month is advised. We believe that the estriol pellets confer no advantages over the usual treatment regimens. We further 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.

Percutaneous Estrogen
Estradiol delivery can be accomplished by the application of a gel to the skin, usually over the upper arms and shoulders or the abdomen and thighs. The preparation produces blood levels of estradiol of approximately 95–125 pg/mL, levels that are both higher and more variable than the standard oral regimens. As with pellets, we recommend that blood estradiol levels be monitored and maintained at a level below 200–200 pg/mL.

The Narrow Therapeutic Indow

There are reasons to believe that the dose of estrogen administered is important in achieving maximal cardiovascular benefits. Hemodynamic responses to estrogen vary according to blood estrogen levels. Relatively normal levels are associated with better left ventricular contraction and performance. Very high estradiol levels achieved with large doses of estrogen produce the opposite effects, a decrease in left ventricular systolic and aortic blood flow. The beneficial effect of postmenopausal estrogen in preventing the hyperinsulinemia associated with aging is present with a dose of 0.625 mg conjugated estrogens, but lost with a dose of 1.25 mg. The effect of estrogen on arterial thrombosis is dose related; we know from our experience with oral contraceptives that high doses of estrogen cause myocardial infarctions and strokes. At least 3 studies have found that estrogen doses greater than 0.625 mg conjugated estrogens are less beneficial in terms of coronary heart disease and mortality; although the numbers of patients on the higher doses were relatively small, and these conclusions did not achieve statistical significance. For these reasons, we believe it is important to avoid excessive dosage. We find it helpful to measure the blood level of estradiol in patients who demand ever-increasing doses of estrogen. Sharing the results of the measurement and the concern regarding high doses helps patients accept the recommendation to keep the blood level of estradiol below 200 pg/mL, and, preferably, below 100 pg/mL.

Is there a dose of estrogen below which the beneficial effects begin to be lost? This is a difficult question to answer because so few women in epidemiologic studies have taken less than the standard doses. In a 2-year study, the lipid effects associated with 0.3 mg esterified estrogen were not significantly different from the placebo group. In the monkey, the cardiovascular beneficial responses begin to diminish below a circulating blood estradiol level of 80 pg/mL. Therefore, it is likely that some reduction in cardiovascular benefit occurs with less than standard estrogen doses.

Studies have demonstrated that a dose of 0.625 mg of conjugated estrogens is necessary to preserve bone density. The conventional wisdom has stated that an estradiol blood level of 40–60 pg/mL, is required to protect against bone loss. 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.

The rate of bone loss and the incidence of hip and vertebral fracture is inversely related to the circulating estrogen levels in older women. 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.

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 1500 mg daily). 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. 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. Major concerns with lower doses include the possibility that there will be a significant percentage of non-responders, and some cardiovascular benefit will be sacrificed. 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 bone response with measurements of either bone density or urinary biochemical markers. After 6 months to one year, we urge patients on lower doses to move up to a standard regimen.

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? For example, a highly specific assay for estradiol will detect the 0.1 pg/mL dose protects the endometrium as well as the 10 mg dose. Unfortunately, progestin withdrawal bleeding occurs in 80–90% of women on a sequential regimen. The sequential regimen can also cause adverse symptoms related to the dose of progestin such as breast tenderness, bloating, fluid retention, and depression. Switching from medroxyprogesterone acetate to norethindrone often relieves these complaints.

The Sequential Program for Oral Postmenopausal Hormone Therapy

<table>
<thead>
<tr>
<th>Daily estrogen:</th>
<th>0.625 mg conjugated estrogens, or 1.25 mg estradiol, or 1.0 mg micronized estradiol or equivalent doses of other estrogens.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly progestin:</td>
<td>0.7 mg norethindrone, or 200 mg micronized progesterone, or 5 mg medroxyprogesterone acetate, or equivalent doses of other progestins given daily for 2 weeks every month.</td>
</tr>
</tbody>
</table>

Combined with daily calcium supplementation (500 mg with a meal), and vitamin D (400 IU in winter months and 800 IU for women over age 70).

In the sequential regimen, the amount of norethindrone equivalent to 10 mg medroxyprogesterone acetate is 1.0 mg. Norethindrone is available in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective. Micronized progesterone is absorbed in a dose of 0.35 mg in the progestin-only minipill oral contraceptive. In a 3-year randomized trial, 200 mg micronized progesterone given daily for 12 days each month effectively protected the endometrium against hyperplasia. The lowest effective dose of micronized progesterone and the proper dose for continuous daily administration are not established; however, a short-term (6 months) study indicated 100 mg/day was effective.

The continuous combined method of treatment evolved to improve patient continuance in the presence of bleeding and other symptoms. The continuous activity of progestins allows the use of lower doses that by virtue of a daily availability inhibit endometrial growth. This approach involves the continuous daily use of the following estrogen-progesterin combinations:
The Continuous Combination Program for Oral Postmenopausal Hormone Therapy

Daily estrogen: 0.625 mg conjugated estrogens, or 1.25 mg estradiopate, or 1.0 mg micronized estradiol, or equivalent doses of other estrogens.

Daily progestin: 0.35 mg norethindrone, or 100 mg micronized progesterone, or 2.5 mg medroxyprogesterone acetate, or equivalent doses of other progestins.

Combined with daily calcium supplementation (500 mg with a meal), and vitamin D (400 IU in winter months and 800 IU for women over age 70).

Continuance with hormone therapy programs is notoriously poor. The two most common reasons why women discontinue or do not start hormone treatment are fear of cancer and vaginal bleeding. The current data on breast cancer are reassuring, and the addition of a progestational agent has effectively prevented endometrial cancer. But the persistence of bleeding with the traditional sequential regimen continues to be a barrier to good continuance. To go from 80–90% withdrawal bleeding to 80% no bleeding represents a major accomplishment, and thus, the continuous approach has a significant advantage.

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 one year. 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: breakthrough bleeding. 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 probably due to a progestational effect on systemic 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 the lower dose (2.5 mg). 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 contact to allay anxiety and encourage persistence.

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 one 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 or 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 in order 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). The local release of progestin is effective in suppressing endometrial response and preventing bleeding, but the progestrone-releasing IUD must be replaced every 18 months. The levonorgestrel-releasing IUD can be left in place for 10 years, a decided advantage. Finally, for some patients, vaginal hysterectomy will prove to be an acceptable alternative.

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. A reasonable economic moderation would be to limit pretreatment biopsies (using the 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. It is prudent to assess the endometrium in these patients prior to changing from unopposed to combined therapy. Clinicians should
A combined estrogen-progestin program will not totally prevent endometrial cancer. 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 has been 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. 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. 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 us with 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 pathologic. However, the severity of pathologic change does not correlate with the measured thickness. Endometrial thickness (the two layers of the anterior and posterior walls in the longitudinal axis) under 5 mm is reassuring and allows conservative management. 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. 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. Because endometrial thickness by ultrasonography in patients on a sequential regimen can be affected by day in the treatment cycle, ultrasonography assessment should be obtained toward the end of the progestin phase or at the beginning of the cycle. 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. Doppler velocimetry does not improve the accuracy of discriminating between normal and abnormal endometrium. 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. 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. 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 that support the efficacy and validity of this method. 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. 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.

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 other than breast tenderness. An occasional patient will develop nausea. This suggests that some of the 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).

Can the progestational agent be administered less frequently? We are secure in our position, supported by clinical data, that a monthly estrogen-progestin sequential or a daily combination program effectively prevents endometrial hyperplasia. Experience with other regimens is very limited. The administration of medroxyprogesterone acetate every 3 months was associated in one study with longer, heavier menses and unscheduled bleeding, and a 1.5% incidence of hyperplasia at one year, while in another study, overall bleeding was less, but the incidence of hyperplasia was approximately 4%. In yet another study, there was no endometrial hyperplasia encountered by 143 women who completed two years of treatment; however, the progestin administered every 3 months was of high dosage, 20 mg medroxyprogesterone acetate daily for 14 days. Most impressively, the Scandinavian LongCycle Study, a clinical trial scheduled to last 5 years, was canceled after 3 years because of a 12% incidence of endometrial hyperplasia and 4 cases of endometrial cancer. Therefore, if a patient chooses this regimen, more intensive endometrial monitoring is required. In our view, an annual endometrial biopsy is strongly recommended. Indeed, 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.

Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate. Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate. Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate. Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate. Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate. Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to chlormadinone acetate or cyproterone acetate.

Progestins Available Worldwide

<table>
<thead>
<tr>
<th>Progestins Available Worldwide</th>
<th>Estimated</th>
<th>Comparable Doses</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-Carbon derivatives.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medroxyprogesterone acetate</td>
<td>5.0 mg</td>
<td></td>
</tr>
<tr>
<td>Megestrol acetate</td>
<td>5.0 mg</td>
<td></td>
</tr>
<tr>
<td>Cyproterone acetate</td>
<td>1.0 mg</td>
<td></td>
</tr>
<tr>
<td>Dydrogesterone</td>
<td>10.0 mg</td>
<td></td>
</tr>
<tr>
<td>Chlomadinone acetate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Medrogestone
Demegestone
Promegestone
Trimegestonel
Nomegestrol acetate
19-Nortestosterone Family:
Norethindrone 0.7 mg
Norethindrone acetate 1.0 mg
Levonorgestrel 0.75 mg
Desogestrel 0.15 mg
Nomegestrol
Norethynodrel
Lynestrenol
Ethynodiol diacetate
Gestodene
Norgestimate
 Dienogest

The Progestin Intrauterine Device

The contraceptive levonorgestrel-releasing intrauterine device (IUD) has been reconfigured in a smaller model releasing as little as 5 µg of levonorgestrel per 24 hours. It has been reported that patients who have had Stage I or Stage II adenocarcinoma of the endometrium can take estrogen without fear of recurrence, and combined regimens, there is irregular breakthrough bleeding in the first 6 months, and after one year, approximately 60–70% of the women are amenorrheic. The levonorgestrel device has the advantage of a 10-year duration of use. The progesterone-releasing device is larger, must be replaced every 18 months, and there have been few studies in postmenopausal women.

Should Progestins Be Administered to 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 being treated with unopposed estrogen, the combined estrogen-progestin program is strongly advised in patients with a past history of endometriosis. In addition, we have encountered a case of hydroureter and 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 had 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, and combined estrogen-progestin treatment is recommended for these women.

3. It has been reported that patients who have had Stage I or Stage II adenocarcinoma of the endometrium can take estrogen without fear of recurrence. Careful studies indicate that the addition of medroxyprogesterone provides an additional effect on bone only in women with established, significant osteoporosis.

4. The combined estrogen-progestin approach makes sense for patients previously treated for endometrioid tumors of the ovary.

5. There is evidence that the combination of estrogen and progestin has a greater positive impact on bone density than estrogen alone. Thus, in hysterectomized women at high risk for osteoporosis, the combined estrogen-progestin program may offer an important potential advantage. However, this synergistic effect is influenced by the type of progestin, and a greater bone response is probably limited to the 19-nortestosterone (norethindrone) family.

6. In women with elevated triglyceride levels, the addition of a progestin, especially a 19-nortestosterone progestin, may attenuate a further estrogen-induced increase in triglycerides.

The Addition of Androgens

After menopause, the circulating level of androstenedione is about one half that seen prior to menopause. Most of this postmenopausal androstenedione is derived from the adrenal gland, with only a small amount secreted from the ovary. Testosterone levels do not fall appreciably, and in most women, the postmenopausal ovary, for a few years, actually secretes more testosterone than the premenopausal ovary. The remaining active stromal tissue in the ovary is stimulated by the elevated gonadotropins to this level of increased testosterone secretion. The total amount of testosterone produced, 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. These effects, however, have been reported with the administration of relatively large doses of androgen. 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.

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. 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. In addition, 30% of the patients experienced acne, and 36% developed facial hirsutism. A lower combination of this (0.625 mg esterified estrogens and 1.25 mg methyltestosterone) also significantly lowers HDL-cholesterol. The adverse impact on the lipid profile is less (and may even be avoided) by the parenteral administration of testosterone. Of course, the clinical effects of these metabolic changes are not known.

It should also 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 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.
estrogen therapy program has been reported to provide no additional beneficial impact on bone or on relief from hot flushes.\textsuperscript{12, 13} 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.\textsuperscript{14} 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.\textsuperscript{15} 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 where the patient and clinician are convinced 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. After some months or a few years, conversion to a standard program without androgen is recommended.

**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 agents with progressively better agonist/antagonist profiles, yielding increasingly user friendly drugs.

Raloxifene exerts no proliferative effect on the endometrium but produces favorable responses in bone and lipids.\textsuperscript{16, 17} Early results (2–3 years of treatment) indicate that women receiving raloxifene have a reduction in the incidence of estrogen receptor-positive breast cancer.\textsuperscript{15} However, because malignant breast tumors require a relatively long period of time to progress from an abnormal cell to a clinically detectable mass, the impact of a specific drug treatment may reflect growth acceleration (as discussed with estrogen therapy) or deceleration of a pre-existing tumor, not causation or prevention.

Short-term clinical trial data indicate that raloxifene has a positive impact on bone density, favorable effects on LDL-cholesterol, fibrinogen, and lipoprotein(a), but no effects on HDL-cholesterol and PAI-1, and raloxifene increases hot flushing.\textsuperscript{18, 19} Long-term clinical trial data will be necessary to determine the ultimate impact on clinical events, specifically fractures, coronary heart disease, stress incontinence, endometrial cancer, and cognition. The MORE study (Multiple Outcomes of Raloxifene Evaluation) is a worldwide clinical trial of raloxifene in osteoporotic women. After 2 years, raloxifene treatment was associated with a 44% reduction in vertebral fractures.\textsuperscript{20}

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.\textsuperscript{21} In a rabbit model, raloxifene did inhibit aortic atherosclerosis, but not as effectively as estrogen treatment.\textsuperscript{22} However, a combination of actions (androgens and estrogen) with some beneficial effects on lipids, a reduction in homocysteine levels) makes it likely there will be some favorable impact on the cardiovascular system.\textsuperscript{23} RUTH (Raloxifene Use for the Heart) is both a primary and secondary coronary heart disease international prevention trial, started in 1998, that should provide data by 2003–05.

At the present time, in our view, raloxifene is an option for prevention of osteoporosis, especially for patients reluctant to use hormone therapy, but not a substitute for estrogen. Remember that women can develop hot flushing on raloxifene, and a 3-fold increase in leg cramping is a problem that is thus far unique for raloxifene.

**Does Hormone Therapy Cause Fibroid Tumors to Grow?**

Uterine leiomyomas are monoclonal tumors that retain sensitivity to both estrogen and progesterin (Chapter 4) and, therefore, it is appropriate to be concerned over whether leiomyoma 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-progesterin therapy.\textsuperscript{24} However, the hormonal dose in this study was relatively high, certainly higher than standard regimens. At the end of one 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 comparison to an increase in size observed with transdermal estradiol (50 µg) and 5 mg medroxyprogesterone acetate daily (a response that may reflect the effect of a higher progestin dose).\textsuperscript{25} In follow-up studies with standard doses of estrogen-progesterin or tibolone, ultrasonography has detected no changes in uterine or myoma volumes.\textsuperscript{26} Clinical experience indicates that fibroid tumors of the uterus almost always are not stimulated to grow by the usual postmenopausal doses of estrogen and progesterin. Nevertheless, pelvic examination and ultrasound surveillance is a wise course. For example, a vulvar leiomyoma with growth stimulated by estrogen-progesterin treatment has been reported.\textsuperscript{27} A case-control study could find no statistically significant increase in the risk of uterine sarcomas associated with estrogen therapy.\textsuperscript{28}

**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 a careful study, however, postmenopausal hormone therapy had no significant effect on sleep disordered breathing in women with more than mild obstructive sleep apnea.\textsuperscript{29} The slight rise in basal body temperature induced by a prostaglandin 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 Rheumatic Diseases**

No clear connection 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, while other studies find no effect.\textsuperscript{30} These studies have been hampered by small numbers, a problem that will not be overcome unless postmenopausal hormone therapy becomes more widespread. 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.\textsuperscript{31} 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 twofold increase in systemic lupus erythematosus, an observation based on 30 cases in past and current users of estrogen.\textsuperscript{32} If this epidemiologic association is true, the absolute risk is very small, and, importantly, postmenopausal women with systemic lupus erythematosus may derive substantial benefit from the cardiovascular actions of estrogen. In a follow-up of 60 postmenopausal women with stable systemic lupus erythematosus, no adverse effects of hormone therapy could be detected.\textsuperscript{33} Furthermore, patients with systemic lupus erythematosus treated with glucocorticoids are at greater risk for osteoporosis.\textsuperscript{34} Nevertheless, there is a concern that exogenous estrogen will increase flares and stimulate activity, some beneficial effects on lipids, a reduction in homocysteine levels) makes it likely there will be some favorable impact on the cardiovascular system.\textsuperscript{35} SELENA (Safety of Estrogen in Lupus Erythematosus National Assessment) is an ongoing randomized, controlled clinical trial of hormone therapy and lupus erythematosus in postmenopausal women, as well as oral contraceptive therapy in premenopausal women.

Bone loss associated with glucocorticoid therapy can be avoided with the usual postmenopausal hormone regimens.\textsuperscript{36} These patients are also excellent candidates for bisphosphonate treatment, another effective option that prevents glucocorticoid-induced bone loss.\textsuperscript{37} In addition, calcium and vitamin D supplementation has been demonstrated to prevent bone loss associated with low-dose glucocorticoid treatment.\textsuperscript{38}

**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,\textsuperscript{39} and, 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.\textsuperscript{40} However, 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.\textsuperscript{41} Because there are no known treatments that modify the course of arthritis, this potential benefit of postmenopausal hormone therapy deserves study by a randomized clinical trial.
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 could be 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. 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. 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. Prevention with hormone therapy or bisphosphonate treatment should be considered.

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. Estrogen therapy is also associated with a reduction in gingival inflammation and bleeding. 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 should also be manifested on oral bone. 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. Approximately a 25% reduced risk of tooth loss in current users of estrogen has been observed in the Nurses' Health Study.

Professional singers have used hormone therapy to prevent what they view as unwanted voice changes associated with menopause. In a 1-year study, objective voice analyses documented a more androgenic change in voice in the early postmenopausal years with a lesser change associated with estrogen treatment, slightly attenuated by the addition of a progestin.

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. An increased prevalence of keratoconjunctivitis sicca (dry eyes) in menopausal and postmenopausal women is recognized by ophthalmologists, and estrogen therapy offers the potential for symptomatic relief. There is further evidence that postmenopausal estrogen therapy has an effect that protects against lens opacities. Estrogen-progestin treatment also lowers intraocular pressure in postmenopausal women with normal eyes or glaucoma.

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. This association between femoral neck bone mass and age-related hearing loss suggests that prevention of bone loss with estrogen therapy might also prevent to some degree age-related hearing loss. It is possible that this potential action of estrogen could be exerted on the cochlear capsule.

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 over age 65. This is a strong argument in favor of treating very old women who have never been on estrogen. Estrogen treatment that is not begun until after age 60 can with long-term use achieve bone densities nearly 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. Whether estrogen's cardiovascular protection has anything significant to contribute in the elderly has not been addressed. It makes clinical sense, however, that some positive contribution can be expected. 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.

If postmenopausal hormone therapy is demonstrated to have a beneficial effect on the risk and severity of Alzheimer's disease, this would become a powerful reason to recommend treatment for very old 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. It is usually better to start older women with lower doses. This can utilize 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. After 6 months, an increase to standard doses is recommended to maximize the bone, cardiovascular, and central nervous system benefits.

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 5th year after discontinuing treatment.

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. Flavonoids (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. This raises the possibility that dietary interactions with food products could produce a clinical impact. It is well recognized that 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. However, when alcohol is administered under experimental conditions, circulating estrogen concentrations are raised to high levels. 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.

Alternative Treatments for Hot Flushes

When patients with hot flushing cannot take estrogen, we recommend transdermal clonidine, applied with the 100 μg dose once weekly. Side effects are minimal, and a modest impact can be expected.

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.
Tibolone

Tibolone is a steroid, related to the 19-nortestosterone family, that is effective for the treatment of bone and hot flushes in a dose of 2.5 mg per day. It is metabolized into 3 steroid isomers with varying estrogenic, progestogenic, and androgenic properties. The metabolites differ in their activities and dominance according to the target tissue. Thus, tibolone provides estrogenic effects on bone and hot flushing, but it induces atrophy of the endometrium.

Beneficial impact on bone (2.5 mg dose) is comparable with standard hormonal therapy. A lower dose of 1.25 mg daily also provides bone protection, but it is less effective and there is more vaginal bleeding. In the endometrium, tibolone is converted locally (by endometrial 3 beta-hydroxysteroid dehydrogenase/isomerase) to its D19 progesterational isomer; hence, tibolone exerts a progestational effect on the endometrium. Tibolone has an estrogenic effect on the vagina, and women report improvements in the symptoms of vaginal dryness and dyspareunia, and an increase in sexual enjoyment and libido.

Although a short-term reduction in HDL-cholesterol is an undesirable consequence with tibolone treatment; the long-term impact on the risk of cardiovascular disease is unknown. In a 2-year study, the unfavorable effect on lipoproteins was accompanied by beneficial changes in coagulation factors consistent with enhanced fibrinolysis and unchanged coagulation. Overall, it is possible that some favorable activity on the cardiovascular system is maintained. A major advantage of tibolone (2.5 mg daily) is its low (10-20%) incidence of bleeding. Because tibolone inhibits breast cell proliferation in vitro, it is possible that future studies will indicate that tibolone offers some protection against breast cancer. Tibolone also has a beneficial impact in short-term studies on insulin resistance in normal women and in women with non-insulin-dependent diabetes mellitus.

"Natural" Therapies

Patients should be questioned regarding the use of "natural" therapies. Some of the herbs that contain estrogen-like compounds include ginseng, agnus castus, red sage, black cohosh, and beth root. The dosage and purity of herbal preparations are unknown, and most importantly, there are no substantial studies documenting either harmful or beneficial effects. Herbs are often contaminated with heavy metals. A rigorous evaluation of one popular herb, dong quai, could detect no effects on vaginal maturation or menopausal symptoms, especially hot flushing. In view of our experience, the use of products without scientific study should be discouraged. We believe 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.

Phytoestrogens

Phytoestrogens are classified into 3 groups: isoflavones, coumestans, and lignans. They are present in many 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 glycitein. These phytoestrogens are characterized by mixed estrogenic and antiestrogenic actions, depending on the target tissue. Variations in activity may also be due to the fact that the soy phytoestrogens have a greater affinity for the estrogen receptor-beta compared with estrogen receptor-alpha. Estradiol and soy protein produce comparable metabolic changes in the monkey, including favorable lipid changes, improved carbohydrate metabolism, and a decrease in central, androgen abdominal fat. Human clinical trials have resulted in inconsistent lipid effects. The soybean phytoestrogens do not maintain bone density in the monkey, but do have favorable effects on atherosclerosis and vasomotor responses, although the effect on atherosclerosis is not as robust as that of estrogen.

The daily intake of dietary soy reduces the number of hot flushes in postmenopausal women, although there is significant variability in response, and efficacy appears to be less than that of estrogen. In women, soy consumption produces a reduction in the circulating levels of estradiol, and it is suggested that the replacement of estradiol target with specific phytoestrogens may be beneficial. 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 another case-control study indicated a reduction in the risk of breast cancer, in women with a high consumption of soy and other legumes. It is by no means certain, however, that there is a direct effect of soy intake. Indeed, a 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. It will require appropriate clinical trials to determine how phytoestrogens compare to estrogens, and 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, handicapped by gastrointestinal symptoms, a major alteration in diet or the use of an unacceptable supplement, and great variability in plant contents and products. In addition, individuals demonstrate great variability in absorption and metabolism. A user-friendly preparation needs to be developed that minimizes individual variability in response.

Ipriflavone is a synthetic product with a structure that is very similar to the naturally occurring isoflavones. It prevents postmenopausal bone loss with no major side effect problems. Larger clinical trials are required to determine the variability in response.

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. 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. Animal studies (in animals that do not even synthesize DHEA) have suggested that DHEA administration enhances the immune system and protects against many conditions associated with aging. Preliminary human studies have indicated significant increases (even at the usual 50 mg/day dose) in testosterone and dihydrotestosterone levels in women (by peripheral conversion in a variety of tissues). Short-term studies of even small doses (25 mg/day) indicate that DHEA administration produces an adverse effect on the lipid profile. What is the impact of a long-term increase in androgens? It is prudent to await the outcome of clinical trials before advocating DHEA supplementation.

Miscellaneous Considerations

A striking and consistent finding in most studies dealing with menopause and hormonal therapy is a marked placebo response in a variety of symptoms, including flushing. A significant clinical problem encountered in our referral practice is the following scenario: a woman will occasionally undergo an apparent 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 therapy. 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.

Assessing vaginal cytology is not useful. The vaginal mucosa is too sensitive to estrogen to allow dose-response titering.

High-dose calcium supplementation can unmask asymptomatic hyperparathyroidism. Women receiving calcium supplementation in excess of 500 mg daily should have their blood levels of calcium and phosphorous measured yearly for the first 2 years. If normal, no further surveillance is necessary.

Assessment of the cholesterol-lipoprotein profile should follow the guidelines for general preventive care, with one exception. No further measurements are required in women on postmenopausal hormone therapy, neither before nor during treatment. The one exception is the patient with previous evidence of elevated triglycerides. After 2–4 weeks of hormone treatment, the triglyceride level should be measured to make sure the patient is not the rare individual who responds to estrogen with excessively high triglyceride levels.

The Problems of Estrogen-Progestin Therapy

Metabolic

Patients with high-risk factors need special consideration when estrogen therapy is being considered. Metabolic contraindications to estrogen therapy include: chronically impaired liver function, acute vascular thrombosis (with or without embol), and neuropathologic vascular disease. Estrogens may have adverse effects on some patients with seizure disorders, familial hyperlipidemias (very high triglycerides), and migraine headaches.

Pancreatitis and severe hypertriglyceridemia can be precipitated by the administration of oral estrogen to women with elevated triglyceride levels. In women with triglyceride levels between 250 mg/dL and 750 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 750 mg/dL represents an absolute contraindication to estrogen treatment. Although triglyceride levels in the normal range were not affected by progestins in the PEPI trial (see Chapter 17), 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 should be considered for women with elevated triglycerides.

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. The risk of cholecystectomy appeared to increase with dose and duration of use and to persist for 5 or more years after stopping treatment. Others have also reported increased risks of cholecystectomy in past and current users of estrogen. In the HERS clinical trial, the relative risk of gallbladder disease was 1.38; however, this did not achieve statistical significance. 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. A cross-sectional study of gallstone disease could detect no association with postmenopausal hormone treatment.

The routine, periodic use of blood chemistry is not cost-effective, and careful monitoring for the appearance of the symptoms and signs of biliary tract disease will suffice. It is not certain that this potential problem is limited to oral therapy. Non-oral routes of estrogen administration have been reported to both increase and not increase biliary cholesterol saturation (a lithogenic response).

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. Weight gain in women at menopause is not due to the hormonal changes associated with menopause. 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. In fact, in the PEPI trial, the hormone-treated groups actually gained less weight than the placebo group.

Estrogen (with or without progestin) prevents the tendency to increase central body fat with aging. This would inhibit the interaction among abdominal adiposity, hormones, insulin resistance, hyperinsulinemia, blood pressure, and an atherogenic lipid profile.

Venous Thrombosis

It is well recognized that pharmacologic doses of estrogen (oral contraceptives) are associated with an increased risk of venous thrombosis. The impact of the lower doses administered to postmenopausal women has been more difficult to ascertain. Older case-control studies failed to find a link between postmenopausal doses of estrogen and venous thrombosis. However, these studies excluded cases with pre-existing risk factors for thrombosis. A well-designed case-control study of older women unselected for other thrombolic risk factors indicated that postmenopausal doses of estrogen did not increase the risk of venous thrombosis. Others have also failed to find an increase in venous thromboembolism associated with postmenopausal hormone therapy.

The conventional wisdom that the low postmenopausal doses of estrogen do not increase the risk of venous thrombosis was challenged more recently by 4 case-control studies and one cohort study. These studies were larger, but still were limited to 20–40 cases on hormone therapy. In the Nurses’ Health Study cohort, the risk of pulmonary embolism was increased 2-fold in the current hormone users. The case-control studies found a 2.1- to 3.6-fold increased risk of deep vein thrombosis in current but not past hormone users. The case numbers were too small to enable a reliable analysis of a dose-response effect or the effect of transdermal use. The studies all indicated that the increased risk was confined to early use, and that the risk lowers to a non-significant level after one year. There was no indication that the addition of a progestin changed the risk.

The overall impact of postmenopausal hormone therapy on the clotting cascade is compatible with a favorable effect due to changes consistent with an increase in fibrinolysis. However, the favorable impact may be confined to arterial thrombosis. Furthermore, it is always more prudent to make clinical judgments based on epidemiologic event data rather than laboratory tests. However, epidemiologic study of venous thrombosis (deep vein thrombosis and pulmonary embolism) is difficult because of the low incidence of the complication and the problem of diagnostic bias (a clinician is more suspicious of the diagnosis because of the use of estrogen). 

The clinching argument on this issue, in our view, is the observation of an increased risk of venous thrombosis in both the HERS secondary prevention study of daily estrogen-progestin therapy for women who already have coronary heart disease and the data obtained with rafoxane. The HERS (Heart and Estrogen-Progestin Replacement Study) recorded a 2.89 relative risk of venous thromboembolism comparable with that reported in the observational studies. The increased risks in these instances were observed in randomized clinical trials, and it is difficult, if not impossible, to not accept these results.

What is the final message for clinicians and patients? First, it should be emphasized that the risk appears to apply only to new starters; women who have been on hormone therapy can be reassured that there is no evidence of an increased risk of venous thrombosis after the initial one year of treatment. The actual risk is very low because of the low frequency of this event. If the relative risk is increased 3-fold, this would increase the incidence of venous thromboembolism by about two 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.

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. The following recommendations 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.
Thrombophilia Screening

- Antithrombin III deficiency
- Protein C deficiency
- Protein S deficiency
- Factor V Leiden mutation
- Prothrombin gene mutation
- Antiphospholipid syndrome

A DNA-based test can be used to verify the presence of the factor V Leiden mutation. 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, and unlike arterial thrombosis, smoking either has no effect or at best is a weak risk factor for venous thromboembolism.

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 and chronic anticoagulation, in consultation with a hematologist.

There are no studies of venous thromboembolism following surgical procedures in postmenopausal hormone users. It now seems sensible to recommend appropriate prophylactic anticoagulant treatment in hormone users anticipating immobility with hospitalization, especially if other risk factors 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 empirical, individual decision. Hormone therapy can be resumed when the patient is ambulatory.

Endometrial Neoplasia

There are two different types of endometrial cancer. The more uncommon form (perhaps 20%) develops rapidly, usually in older women, with a histologic pattern more characteristic of serous carcinomas, in a background of atrophic endometrium. 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 one 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 PEP1 trial, 30% of the women on unopposed estrogen developed adenomatous or atypical hyperplasia. 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.

Approximately 40 case-control and cohort studies have 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 1000 postmenopausal women per year. 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, and perhaps an incidence of 1 in 10 with very long-term use), lingers for up to 15 years after estrogen is discontinued, and 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. Although most endometrial cancer associated with estrogen use is of low grade and stage, and associated with better survival (probably because of earlier 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).

A short-term study (2 years) has indicated that 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. We have learned that long-term exposure to low levels of estrogen can induce abnormal endometrial growth, 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 contained 18 cases and 9 controls who had exclusively used only 0.3 mg/day of unopposed conjugated estrogens. 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.

It is now apparent that this risk can be reduced by the addition of a progestational agent to the program. 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 an 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. 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. While one standard method incorporated the addition of a progestational agent for the last 10 days of estrogen exposure, most have argued in favor of 12 or 14 days. Studies indicate that the minimal requirement is a monthly exposure of at least 10 days' duration. About 2–3% of women per year develop endometrial hyperplasia when the progestin is administered for less than 10 days monthly.

Important unanswered questions are what will be the actual incidence of endometrial cancer in long-term users of postmenopausal hormone therapy and will there be 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. 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 2 deaths, precluding statistical significance. 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.

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. It is just as logical to believe that prevention of growth and development of atrophic endometrium are protective. There is good reason to believe that both the sequential regimens (with appropriate dose and duration of progestin administration) and the continuous combined regimen offer protection against endometrial cancer. The degree of protection and comparable performance will ultimately be determined by the long-term randomized clinical trials currently ongoing.

The lowest daily dose of progestin that protects the endometrium has not been established. Currently, the sequential program utilizes 5 or 10 mg medroxyprogesterone acetate and the combined daily method uses 2.5 mg. The dose of norethindrone that is comparable with 2.5 mg medroxyprogesterone acetate is 0.25 mg. Although lower doses of progestational agents are effective in achieving target tissue responses (such as reducing the nuclear concentration of estrogen receptors), the long-term impact on endometrial histology has not yet been firmly established. The question of dose is an issue of major importance, especially in terms of the cardiovascular system and compliance because of progestin-induced side effects.
Although the protective effect of progesterin is considerable and predictable, it is unwise to expect all patients on estrogen-progesterin 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.

Ovarian Cancer

A prospective cohort study concluded that the risk of fatal ovarian cancer is increased with long-term estrogen use. There was no significant increase in the relative risk of fatal ovarian cancer with ever use of postmenopausal estrogen, and the link with long-term estrogen use achieved statistical significance only with the 18 cases using estrogen for 11 or more years. By no means is it certain if this association is real. There have been 12 case-control studies of ovarian cancer risk factors, and the pooled analysis of this literature could find no consistent evidence for an association between ovarian cancer and estrogen therapy. 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. We are not impressed by this small increase in a meta-analysis of case-control studies, subject to multiple potential biases. Among the 6 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.

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 more recent and 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. Another case-control study reported a slightly increased risk, but it was not statistically significant. In a retrospective analysis, no detrimental impact could be detected on prognosis after surgery for ovarian cancer in patients who received postmenopausal hormone therapy after diagnosis.

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. 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). 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.

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. 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 35% reduced risk in current users) could not demonstrate an added benefit with longer duration of current use. A reduction in colonic adenoma has been documented in current users. There also appears to be a reduced risk of polyps, especially large polyps, among current and recent hormone users.

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. This potential benefit deserves greater attention; colorectal cancer ranks third in women, both in incidence and mortality, more prevalent than cancers of the uterus or ovary.

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. 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 another case-control study could find no association with postmenopausal estrogen treatment. 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.

In an analysis of cancer incidence in a Swedish cohort of women prescribed postmenopausal hormone therapy, no increase in malignant melanoma was observed.

Breast Cancer

The possibility that estrogen use increases the risk of breast cancer must be intensively scrutinized. The American epidemiologic data on the scope of human female breast cancer are astonishing: 1 of every 8 women will develop breast cancer in her lifetime (assuming an 85-year life expectancy). In America, breast cancer is the leading type of cancer in women (29.7%) and now second to lung cancer as the leading cause of cancer death in women (16.1%), about 10 times the number of deaths from endometrial cancer.

Sufficient evidence exists to indicate the possibility of a slightly increased risk of breast cancer associated with long durations (5 or more years) of postmenopausal estrogen use. However, the epidemiologic data on this relationship are by no means consistent and uniform. A review of the epidemiologic studies on postmenopausal hormone therapy and the risk of breast cancer fails to provide definitive evidence regarding this issue. Nevertheless, we believe that patients must consider this possibility in their informed decision making.

Early studies on estrogen use and breast cancer indicated higher risks in special subcategories, such as women with benign breast disease, long duration of use, or natural versus surgical menopause. A retrospective analysis 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. 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 another case-control study could find no association with postmenopausal estrogen treatment.

Patients with breast cancer are at increased risk of heart attack, stroke, and recurrent breast cancer-related death. The risk of breast cancer recurrence is increased with longer duration of hormone use. The risk of breast cancer is increased with more than 10 years of estrogen use. Estrogen use is associated with a slightly increased risk of colorectal cancer, but the evidence is not strong.

A meta-analysis of hormone therapy and breast cancer risk found no overall increase in risk for hormone therapy users compared to nonusers. However, there was a small increase in risk for hormone therapy users who also used hormone therapy for at least 10 years. Another meta-analysis of hormone therapy and breast cancer risk found a small increase in risk for hormone therapy users compared to nonusers, but this effect was not statistically significant. A meta-analysis of hormone therapy and breast cancer risk found a small increase in risk for hormone therapy users compared to nonusers, but this effect was not statistically significant.

The latest reports from the Nurses’ Health Study represent 16 years of follow-up (1976–1992). During that period, 1935 cases of breast cancer were identified among more than 69,000 postmenopausal women. The analysis revealed that women who had used estrogen in the past (even for 10 or more years) were not at increased risk of breast cancer. However, the relative risk for current users was 1.48 (CI 1.22–1.74) for 5–8 years of use, and 1.46 (CI 1.20–1.78) for 10 or more years.

A Danish case-control study used questionnaires to obtain information on cases and controls and also indicated a slightly increased risk of breast cancer associated with long-term estrogen use. However, the statistical power of this study continues to be limited, and there is no control group in this study. There was no significant increase in the relative risk of breast cancer with more than 10 years of hormone therapy. A meta-analysis of hormone therapy and breast cancer risk found a small increase in risk for hormone therapy users compared to nonusers, but this effect was not statistically significant.

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. We are not impressed by this small increase in a meta-analysis of case-control studies, subject to multiple potential biases. Among the 6 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 conclusion of this study was the indication of increased risk with combined treatment with estrogen-progestin. This conclusion also was not statistically significant. In a retrospective analysis, no detrimental impact could be detected on prognosis after surgery for ovarian cancer in patients who received postmenopausal hormone therapy after diagnosis.

A well-publicized study from Uppsala, Sweden, concluded that estrogen use was associated with a slight increase in the risk of breast cancer (relative risk 1.1). However, the statistical power of this study continues to be limited, and there is no control group in this study. There was no significant increase in the relative risk of breast cancer with more than 10 years of hormone therapy. A meta-analysis of hormone therapy and breast cancer risk found a small increase in risk for hormone therapy users compared to nonusers, but this effect was not statistically significant.

The latest reports from the Nurses’ Health Study represent 16 years of follow-up (1976–1992). During that period, 1935 cases of breast cancer were identified among more than 69,000 postmenopausal women. The analysis revealed that women who had used estrogen in the past (even for 10 or more years) were not at increased risk of breast cancer. However, the relative risk for current users was 1.48 (CI 1.22–1.74) for 5–8 years of use, and 1.46 (CI 1.20–1.78) for 10 or more years.
of use. By virtue of the large numbers in the Nurses' Health Study and the careful analyses by the investigators, reports from this study must be given great credibility. The 16-year follow-up report is disturbing with its finding of an increased risk in current users. Because estrogen users may be examined more frequently, detection bias is a major concern. It is noteworthy that current users had a 14% higher prevalence of mammography compared with never users. Current users were different from never users (history of benign breast disease, birth only once or twice, menarche at age 13 or less, body mass index of 21–23). Another important consideration is the need to adjust for alcohol consumption, an accepted risk factor for breast cancer. In the Iowa Women's Health Study, an increased risk of breast cancer was observed only in those postmenopausal women who consumed one drink or more of alcohol daily. The concern is that alcohol consumption raises estrogen levels in hormone users. The authors concluded that alcohol alone would not explain the observed outcome in the Nurses' Health Study, what is the additive effect of all factors? Thus, the finding of an increased relative risk in long-term current users is not definitive and not free of all confounding variables. The size of the statistical risk is not outside the range of influence by biases.

Based on 359 deaths due to breast cancer, the risk of dying of breast cancer in the Nurses' Health Study was 0.80 (CI 0.60–1.07) for past users, 0.99 (CI 0.66–1.48) for current users with less than 5 years of use, and 1.45 (CI 1.01–2.09) with 5 or more years of use. These mortality data raise a question of "prevalence bias," also called by statisticians as interdependence between the probabilities of disease incidence in a population (an issue of competing risks). Is it possible that the protection against cardiovascular disease is so great with long durations of estrogen use that the long-term current users develop a problem that is prevalent with aging, breast cancer. The never users, deprived of the cardiovascular benefit of estrogen, may develop cardiovascular disease before living long enough to experience breast cancer.

A case-control study from Australia, which attempted to control for secular trends in estrogen use, type of menopause, and duration of estrogen use, concluded that there was no evidence for an association between estrogen use and the risk of breast cancer in postmenopausal women. There is a very helpful study that specifically addressed the relationship between the use of estrogen and benign breast disease. This study is impressive because it is based on 30,366 consecutive breast biopsy specimens with follow-up information on 4227 biopsy specimens in 3303 women (a mean duration of follow-up of 17 years). Analysis indicated that the use of estrogen was associated with a reduced risk of developing breast cancer. Most importantly, in patients with atypical hyperplasia in their biopsies, the use of estrogen did not increase and even lowered the risk of breast cancer. Although the protective effect probably indicates detection/surveillance bias, certainly this is strong evidence that estrogen use does not increase the risk of breast cancer in women with surgically proven benign breast disease, even with atypia.

In a study of 1686 cases and 2077 controls in the eastern U.S., the relative risk for current users was a nonsignificant 1.1 (CI 0.7–1.6), for a duration of use of 15 or more years, the relative risk was 0.9 (CI 0.4–1.9). As in the eastern U.S., a study from Toronto found no evidence for an increased risk in either current or recent users (maximum years, 20). A Canadian meta-analysis of risk of breast cancer associated with long-term estrogen use found that estrogen alone or with progestin-progestin. These large case-control studies failed to support the conclusion of the Nurses' Health Study that current users are associated with an increased risk, even with long durations of use. On the other hand, other studies (of smaller size and limited statistical power) have found an increased risk of breast cancer in current users and long-term users. A prospective study from the National Cancer Institute could document only an increase in the risk of in situ breast cancer (not invasive disease), possibly reflecting surveillance bias in women taking either estrogen alone or a combination of estrogen and progestin.

The Iowa Women's Health Study, like the Nurses' Health Study, is prospectively following a cohort of women (selected in 1985). After 6 years of follow-up, a statistically significant increase in the risk of breast cancer could not be detected in ever users or current users of hormone therapy. A report through 8 years of follow-up focused on whether postmenopausal hormone therapy increased the risks for breast cancer and mortality in women with a family history of breast cancer. Even in women with a positive family history of breast cancer who were current users of hormone therapy for more than 5 years, there was no significant increase in the rate of breast cancer. A very large case-control study found no increased risk of breast cancer associated with the ever use of estrogen alone or estrogen and progestin combinations, and when long-term use for 15 years or more was examined, again no increase in risk was detected.

Meta-Analyses. Meta-analysis is an increasingly popular statistical method in which many studies are combined and undergo rigorous analysis. The term "meta-analysis" was coined in 1976 to indicate the re-analysis of data to answer new questions. The method was first used in social science, and then in the late 1980s, in medicine. Simply put, the purpose of a meta-analysis is to gain the statistical power that is lacking in individual studies.

An Australian meta-analysis of 23 studies of estrogen use and breast cancer concluded "unequivocally" that estrogen use did not alter the risk of breast cancer. In the meta-analysis by Dupont and Page (Nashville, Tennessee), the authors concluded that "considerable and consistent" evidence exists that a daily dose of 0.625 mg conjugated estrogens taken for several years does not appreciably increase the risk of breast cancer. They found no evidence of an association between the duration of treatment and the risk of breast cancer at this dosage. The Copenhagen investigators did not consider family history of breast cancer. They found no evidence of an association between the duration of estrogen use and the risk of breast cancer.

A third meta-analysis was from the CDC. This meta-analysis was conducted by using what the authors called a "dose-response curve" for duration of use. The data for each study was calculated by plotting the breast cancer rate against duration of estrogen use. The authors concluded that an increased risk of breast cancer was present regardless of whether menopause was natural or surgical. The analysis concluded that duration of estrogen use was associated with an increased risk of breast cancer, regardless of whom the estrogen was taken. In this analysis, no increase in risk was noted after 5 years of use, but after 15 years of use, the risk was increased by 30%. The effect was present irrespective of risk factors, such as parity or history of benign breast disease. The effect of estrogen therapy on risk of breast cancer was enhanced in women with a positive family history of breast cancer.

A fourth meta-analysis, from Spain, concluded that estrogen is associated with a very small, but statistically significant, increased relative risk of breast cancer and that the increased risk is higher among current users. Confining their analysis to a dose of 0.625 mg conjugated estrogens; however, the Spanish epidemiologists could not detect a statistically significant increased risk. Indeed, this meta-analysis concluded that an estrogen dose of 0.625 mg conjugated estrogens is safe.

A fifth meta-analysis from the epidemiologists associated with the Nurses' Health Study concluded (based on 25 case-control and 6 cohort studies) that there is no increased risk of breast cancer in ever users of estrogen. Current use was associated with an increased risk (which was lost 2 years after using estrogen), and there was a slight increase with more than 10 years of use in this group but there was no linear trend with increasing duration of use. This observation in long-term users could be influenced by an increased proportion of current users in this group, and the increased risk in current users could be a consequence of detection bias. The statistical power of this meta-analysis was in the ever use category, giving strength to its negative conclusion. This meta-analysis could not detect a link between risk of breast cancer and dosage. Nevertheless, a possible effect of higher doses. The Australia, Nashville, and Spain meta-analyses indicated an increased risk with a daily dose of conjugated estrogens greater than 0.625 mg (or its equivalent).

The Vanderbilt, CDC, and Nurses' Health analyses did not find an enhanced risk in women with a history of benign breast disease. In contrast to the CDC report, the Australian and Nurses' Health analyses found no link between positive family history and estrogen use. The Nashville and Spain investigators did not consider family history.

In a major overview and assessment (yet another meta-analysis) of the world's literature on postmenopausal hormone therapy, commissioned by the American College of Physicians, the authors concluded that long-term use of estrogen was associated with a relative risk of breast cancer of 1.25 (CI 1.04–1.51). This conclusion, in our view, was not appropriately critical and represents a judgment that extends beyond the statistical power available from the many heterogeneic studies. The heterogeneity of the many studies is an important issue: different drugs, different doses, different methods of diagnosis, different comparison and control groups. The American meta-analysis is the only one to raise the question: perhaps the heterogeneity is too great to allow an accurate meta-analysis of this literature.
Deceleration

Tamoxifen (Chapter 16)

Oral contraceptives (Chapter 22)

Pregnancy (Chapter 16)

Return to baseline after therapy

<table>
<thead>
<tr>
<th>Causation</th>
<th>Acceleration and Deceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone therapy</td>
<td>Tamoxifen</td>
</tr>
<tr>
<td>Oral contraceptives</td>
<td>Raloxifene</td>
</tr>
<tr>
<td>Pregnancy</td>
<td></td>
</tr>
</tbody>
</table>
breast cancer at a greater rate.

Doses of estrogen known to protect against osteoporosis and cardiovascular disease (0.625 mg conjugated estrogens and 1.0 mg estradiol) are, at the present time, not known to be associated with any clear-cut increased risk of breast cancer. However, because of the concern raised by some studies and reviews that there is a slightly increased risk of breast cancer associated with long-term use of postmenopausal estrogen, this issue requires discussion during the clinician-patient dialogue regarding postmenopausal hormone therapy.

The comfort found in large numbers with epidemiologic research is lost to us when we must make clinical decisions with individual patients. If we wish to minimize the uncertainty associated with endometrial cancer, we will adopt an appropriate strategy. If we wish to measure possibility or probability of an outcome, we will adopt another. In our view, it is appropriate to emphasize the benefits of postmenopausal hormone therapy, point out the continuing concern regarding the relationship between estrogen use and breast cancer (particularly long-term use), and to emphasize the absence of definitive evidence linking such therapy to an increased risk of breast cancer.

The Risk of Breast Cancer and Estrogen-Progestin Therapy. The addition of a progestational agent to postmenopausal estrogen therapy is now accepted as a standard part of the treatment program. The obvious reason for this combination estrogen-progestin approach is the need to prevent the increased risk of endometrial cancer associated with exposure to unopposed estrogen. Even though this endometrial cancer is not frequently encountered and survival rates are excellent with early disease, the fear of this cancer is a major force in patient continuance, and, therefore, the combined approach is warranted. Clinicians and patients have rapidly turned to the method of a daily combination of estrogen and a progesterin, in order to overcome bleeding which is the second major continuance problem.

Only two reports have claimed that the addition of a progestational agent protects against breast cancer. The first was limited by bias in treatment selection (the breast cancer risk factor profiles were not matched in the treated and untreated groups). The second (the Nachtigall study), although it has been the only randomized, placebo-controlled trial, was hampered by small numbers.

At the present time, the available epidemiologic evidence on the impact of combined estrogen-progestin treatment indicates that neither a protective nor a detrimental effect has yet to be convincingly demonstrated; recent studies find that the addition of a progestin does not change the findings with estrogen alone.

Balancing the information available involving all of the health issues affected by hormone therapy, a combined estrogen-progestin program in appropriate doses continues to offer significant benefits for postmenopausal women. As time goes on, more studies and greater duration of use should provide us with better answers to many of our questions. By virtue of the magnitude of the postmenopausal female population, these questions deserve continuing biologic and epidemiologic research from both the public health and individual points of view.

**SUMMARY: Postmenopausal Hormone Therapy and Breast Cancer**

- Some epidemiologic case-control and cohort studies conclude that long-term (5 or more years) of current use of postmenopausal hormone therapy is associated with a slight increase in the risk of breast cancer, a risk that is less than that associated with postmenopausal obesity or daily alcohol consumption. This conclusion might be due to confounding biases, particularly detection/surveillance bias, and accelerated growth of a pre-existing malignancy.
- Many observational studies have failed to develop evidence that long-term postmenopausal hormone therapy increases the risk of breast cancer.
- All epidemiologic studies fail to find an increased risk of breast cancer associated with short-term (less than 5 years) use or past use of postmenopausal hormone therapy.
- The epidemiologic data indicate that the addition of a progestin to the treatment regimen neither increases nor decreases the risk observed in individual studies.
- The epidemiologic data indicate that a positive family history of breast cancer should not be a contraindication to the use of postmenopausal hormone therapy.
- Women who develop breast cancer while using postmenopausal hormone therapy have a reduced risk of dying from breast cancer. This is because of two factors: (1). Increased surveillance and early detection; and (2). Acceleration of pre-existing tumor growth so that tumors appear at a less virulent and aggressive stage.

**Contraindications to Postmenopausal Hormone Therapy**

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. Nothing is known about the risk in patients with more advanced disease. 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 advised in patients with a past history of endometriosis.

Should a Woman Who Has Had Breast Cancer Use Postmenopausal Hormones?

The increasing incidence of breast cancer, together with earlier detection and treatment, is producing a growing pool of patients for whom the question of estrogen treatment is important and at the same time difficult. The problem is easy to articulate: we have no data. There are absolutely no published clinical trials of sufficient size and scope in which the impact of estrogen treatment has been documented when given to women with previously treated breast cancer.

Because there is good reason to believe that breast cancer is hormonally influenced, it is not hard to understand the breast surgeon or medical oncologist who believes that estrogen treatment is foolish and dangerous. Yet, that position is just as unencumbered by data as the position of the clinician who believes that appropriate patients stand to benefit more from estrogen compared with the unknown risk of breast cancer recurrence.

The single most useful prognostic information in women with operable breast cancer has been the histologic status of the axillary lymph nodes. At 10 years, only 25% of patients with positive nodes are free of disease compared with 75% of patients with negative nodes. If more than 3 nodes are involved, the 10-year survival rate drops to 13%. Because of this recognition for the importance of the axillary nodes, the traditional approach to breast cancer (the Halsted surgical approach) was based on the concept that breast cancer is a disease of stepwise progression. There has been 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 (microscopically disseminated) at the time of presentation. Dissemination of tumor cells has occurred by the time of surgery in many patients, and it is concern over the possible response of these cells that fires the debate over this question: should women who have had breast cancer take postmenopausal hormones?

There are two opposing hypotheses that link breast cancer to a hormonal influence: one assigns the critical role to estrogen; the other, to progesterone.

Is Estrogen the Important Influence? When one reviews the various factors associated with breast cancer, many of the factors support a role for increased exposure to estrogen, specifically the importance of age of menarche, age of menopause, obesity, the effect of pregnancy, and added to this list is the therapeutic effect of an antiestrogen, tamoxifen.

Stanley Korenman, therefore, suggested that the endocrine environment does not cause the disease but influences the susceptibility to carcinogens—the so-called estrogen window hypothesis. The two periods of life when the estrogen window is open (this means long periods of unopposed estrogen) are after menarche before ovulation is established and the perimenopausal years. Estrogen shows a significant increase at about Tanner stage 2, ages 8 to 10. Because ovulation occurs on the average 1.5 years after menarche, this window can last 4 to 5 years. It is well recognized that anovulation is common in the perimenopausal years because of the lesser capabilities of the remaining ovarian follicles. Korenman argued that the open window (estrogen exposure) influences the risk of breast cancer and that the window is closed by exposure to progesterone after ovulation and during pregnancy. This argument is supported by observational studies indicating that anovulatory
of the remaining 40, some will live longer (an average of 2–3 years) because of adjuvant treatment, but only a few. Is the unknown risk with exogenous hormone therapy after breast cancer treatment are underway in Europe; for example, the HABITS study in Sweden.

A clinical U.S. trial is currently under way, providing estrogen to randomized women who have been treated for stage I or II breast cancer, who have been previous treated for breast cancer.

There are reassuring reports that breast cancer survivors using hormone therapy do not experience an obvious increase in recurrent disease. There is a small series of patients who have had breast cancer are aware of the benefits of postmenopausal hormone treatment (especially protection against cardiovascular disease and osteoporosis) and also of the potential risk of 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 of the total U.S. population.

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 the large body of evidence that indicates that breast cancer is a hormone-responsive tumor. The overriding fear of many clinicians (and patients) is that metastatic cells present (perhaps being controlled by various host defense factors) that will be susceptible to stimulation by exogenous hormones.

Because of the current lack of epidemiologic data, both sides of this debate are strongly influenced by theoretical considerations and clinical experiences, which, unfortunately, often become an obstacle to the patient's own informed choice.

There are reassuring reports that breast cancer survivors using hormone therapy do not experience an obvious increase in recurrent disease. There is a small series in which a combination of 0.625 conjugated estrogens and 0.15 mg norgestrel was given continuously for a short period of time (a maximum of 6 months) to women who had been previously treated for breast cancer. Over the next two years, no patients developed recurrence. Another small 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 in a large group of patients with similar characteristics. These patients receiving hormone therapy had the same outcomes when compared with 82 women selected from a cancer registry and not taking hormones. In a report from Australia, 152 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. In a follow-up of 49 women treated with estrogen after treatment for localized breast cancer, only one patient developed recurrent disease. And in another series of 114 women, hormone treatment of disease-free patients was associated with a low rate of recurrence. These patients had both positive and negative nodes and estrogen receptor status. Although it is reassuring that the results conform to an incidence of recurrent disease no greater than expected, the outcomes may reflect biases in clinician and patient decision making that can only be overcome with a proper long-term, randomized clinical trial.

A clinical U.S. trial is currently under way, providing estrogen to randomized women who have been treated for stage I or II breast cancer, who have been disease-free for at least 2 years from estrogen receptor negative disease or 10 years if the estrogen receptor status is unknown. Other randomized trials of hormone therapy after breast cancer treatment are underway in Europe; for example, the HABITS study in Sweden. Because of better treatment and earlier diagnosis, 50–75% of women diagnosed with breast cancer are now cured. Of 100 patients with breast cancer, about 60 will be cured by mastectomy or breast-conserving surgery with radiotherapy and would receive no benefit from adjuvant treatment. Is this group safe for hormone therapy? Of the remaining 40, some will live longer (an average of 2–3 years) because of adjuvant treatment, but only a few. Is the unknown risk with exogenous hormone therapy a valid concern that must be addressed? Although intuitively it seems that the risk-benefit ratio would be more favorable in the presence of negative nodes, negative receptors, and small tumors, is negative estrogen and progesterone receptor assessments sufficient to conclude that the cancer is not sensitive to hormone therapy? And if the...
Exacerbation of pulmonary leiomyomatosis by estrogen therapy.

Psychiatric symptoms in response to estrogen in patients with acute intermittent porphyria.

Unusual anecdotal reports include the following:

To all patients with cervical, ovarian, or vulvar malignancies.

No other cancers (besides those mentioned above) are known to be adversely affected by hormone therapy. Postmenopausal hormone therapy can be administered in the presence of these conditions.

Conditions that do not represent contraindications include controlled hypertension, diabetes mellitus, smoking, and varicose veins. The belief that estrogen is contraindicated in women with seizure disorders, familial hyperlipidemias (elevated triglycerides), and migraine headaches is not well supported.

Close surveillance is indicated for some patients with seizure disorders, familial hyperlipidemias (elevated triglycerides), 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, diabetes mellitus, 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. Estrogens are acceptable in the presence of these conditions.

No other cancers (besides those mentioned above) are known to be adversely affected by hormone therapy.

Postmenopausal hormone therapy can be administered to all patients with 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.
2. Exacerbation of pulmonary leiomyomatosis by estrogen therapy.
4. Idiosyncratic cutaneous symptoms in response to estrogen.
5. An infection with Trichomonas vaginalis resolved after discontinuation of estrogen-prostin therapy. 415

Conclusion
We hope 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 to obtain preventive health care benefits. We have learned this from the women who revealed what they believe and what they know in the longitudinal
studies of the last decade. 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, w 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, regular exercise, and control of body weight.
Chapter References
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.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43.
44.
45.
46.
47.
48.
49.
50.
51.
52.
53.
54.
55.
56.
57.
58.
59.
60.
61.
62.

Fosbery WHS, Severe climacteric flushings successfully treated by ovarian extract, Br Med J i:1039, 1897.
Riis BJ, Hansen MA, Jensen AM, Overgaard K, Christiansen C, Low bone mass and fast rate of bone loss at menopause: equal risk factors for future fracture: a 15-year follow-up study,
Bone 19:9, 1996.
after 16 years of age at the levels of lumbar spine and femoral neck in female subjects, J Clin Endocrinol Metab 75:1060, 1992.
Genant HK, Cann CE, Ettinger B, Gordan GS, Quantitative computed tomography of vertebral spongiosa: a sensitive method for detecting early bone loss after oophorectomy, Ann Intern
Field CS, Ory SJ, Wahner HW, Herrmann RR, Judd HL, Riggs BL, Preventive effects of transdermal 17b-estradiol on osteoporotic changes after surgical menopause: a two-year
Ross D, Rees M, Godfree V, Cooper A, Hart D, Kingsland C, Whitehead M, Randomised crossover comparison of skin irritation with two transdermal oestradiol patches, Br Med J 315:288,
1997.
Endocrinol Metab 70:1568, 1990.
Hillard TC, Whicroft SJ, Marsh MS, Ellerington MC, Lees B, Whitehead MI, Stevenson JC, Long-term effects of transdermal and oral hormone replacement therapy on postmenopausal
Lufkin EG, Wahner HW, O'Fallon WM, Hodgson SF, Kotowicz MA, Lane AW, Judd HL, Caplan RH, Riggs BL, Treatment of postmenopausal osteoporosis with transdermal estrogen, Ann
Michaësson K, Baron JA, Farahmand BY, Johnell O, Magnusson C, Persson P-G, Persson I, Ljunghall S, on behalf of the Swedish Hip Fracture Study Group, Hormone replacement
with mild hypertension, Am J Cardiol 78:1385, 1996.
Pines A, Fisman EZ, Drory Y, Shapira I, Averbuch M, Eckstein N, Motro M, Levo Y, Ayalon D, The effects of sublingual estradiol on left ventricular action at rest and exercise in
Stampfer MJ, Colditz GA, Willett WC, Manson JE, Rosner B, Speizer FE, Hennekens CH, Postmenopausal estrogen therapy and cardiovascular disease: ten-year follow-up from the
Williams JK, Honoré EK, Washburn SA, Clarkson TB, Effects of hormone replacement therapy on reactivity of atherosclerotic coronary arteries in cynomolgus monkeys, J Am Coll Cardiol
Ettinger B, Pressman A, Sklarin P, Bauer DC, Cauley JA, Cummings SR, Associations between low levels of serum estradiol, bone density, and fractures among elderly women: The Study
Mizunuma H, Okano H, Soda M, Kagami I, Miyamoto S, Tokizawa T, Honjo S, Ibuki Y, Prevention of postmenopausal bone loss with minimal uterine bleeding using low dose continuous
Stevenson JC, Cust MP, Gangar KF, Hillard TC, Lees B, Whitehead MI, Effects of transdermal versus oral hormone replacement therapy on bone density in spine and proximal femur in
Woodruff JD, Pickar JH, for The Menopause Study Group, Incidence of endometrial hyperplasia in postmenopausal women taking conjugated estrogens (Premarin) with
Strickland DM, Hammond TL, Postmenopausal estrogen replacement in a large gynecologic practice, Am J
Archer DF, Pickar JH, Bottiglioni F, for The Menopause Study Group, Bleeding patterns in postmenopausal women taking continuous combined or sequential regimens of conjugated
The Writing Group for the PEPI Trial, Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: the Postmenopausal Estrogen/Progestin




Because approximately 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. About 25% of U.S. women are overweight, and an additional 25% are obese. Unfortunately, for over 100 years the incidence of obesity has been increasing in the United States and Europe, a reflection of an affluent society with an increasingly sedentary life combined with high caloric foods. This change in lifestyle has produced an increasing prevalence of obesity that is similar in adults and children.

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. 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 Nurses' Health Study estimates that 23% of all deaths in nonsmoking middle-aged women are attributable to being overweight.

The incidence of hypertension, heart disease, noninsulin-dependent diabetes mellitus, gout, gallbladder disease, colorectal cancer, endometrial cancer, postmenopausal breast cancer, and osteoarthritis is elevated in overweight people. Being overweight in adolescence is even a more powerful predictor of cardiovascular adverse health effects than being overweight as an adult.

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 are not the simple answers.

Definition of Obesity

Obesity is an excess storage of triglycerides in adipose cells. There is a difference between obesity and overweight. Obesity is an excess of body fat. Overweight is a body weight 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 weight in pounds:

Women: 100 + (4 x (height in inches minus 60))

Men: 120 + (4 x (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 has been found to correspond closely to densitometry measurements.

The body mass index (the Quetelet index) is the ratio of weight divided by the height squared (in metric units):

\[ \text{BMI} = \frac{\text{weight in kg}}{\text{height in m}^2} \]

To use the nomogram for body mass index (BMI), read the central scale by aligning a straightedge between height and body weight. Over the last 30 years, the mean body mass index has increased in the U.S., and now averages 26.3. A body mass index of 27 or more warrants treatment. A body mass index of about 30 is roughly equivalent to 30% excess body weight, the point at which excess mortality begins (approximately 10–12% of people in the U.S. have a body mass index of 30 or greater). Above age 40, the risk from obesity itself is comparable to that associated with major health problems such as hypertension and heavy smoking. A "good" BMI for most people is in the range of 20 to 24. Mortality is minimal although still increased at a body mass index of approximately 22 and lowest in middle-aged women with a body mass index below 19.
A person is obese when the amount of adipose tissue is sufficiently high (20% or more over ideal weight) to detrimentally alter biochemical and physiological 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 has documented a continuing correlation between the body mass index and cardiovascular disease, diabetes, and cancer. 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.

13 The Nurses’ Health Study has documented a continuing correlation between the body mass index and cardiovascular disease, diabetes, and cancer. 7,14,15 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. 17 The difference between men and women is even greater in older age.

18 Unfortunately, the basal metabolic rate decreases with age. 18,19 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. 20 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 per gram of triglyceride compared to 1 calorie per gram 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 caloric 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. 21

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 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.

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 will, therefore, be low, and glucose will be 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 demonstrate a positive correlation between fewer meals and a greater tendency toward obesity. The person who does not eat all day and then stocks up at night is perhaps doing the worst possible thing.

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 safety 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. 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 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. Neuropeptides that inhibit appetite include corticotropin-releasing hormone (CRH), neuropeptide Y, and cyclo(HisPro), a peptide derived by proteolysis of thyrotropin-releasing hormone. Although recent attention has focused on leptin and the ob gene, keep in mind that 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, “leptoς,” 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. Overeating and obesity, however, are not due to ventromedial nucleus damage but rather to destruction of the nearby ventral noradrenergic bundle. 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 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. Neuropeptides that inhibit appetite include corticotropin-releasing hormone (CRH), neuropeptide Y, and cyclo(HisPro), a peptide derived by proteolysis of thyrotropin-releasing hormone. Although recent attention has focused on leptin and the ob gene, keep in mind that 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, “leptoς,” 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. Overeating and obesity, however, are not due to ventromedial nucleus damage but rather to destruction of the nearby ventral noradrenergic bundle. 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.

The word leptin is derived from the Greek word, “leptoς,” 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. Overeating and obesity, however, are not due to ventromedial nucleus damage but rather to destruction of the nearby ventral noradrenergic bundle. 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.

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 safety 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. 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 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. Neuropeptides that inhibit appetite include corticotropin-releasing hormone (CRH), neuropeptide Y, and cyclo(HisPro), a peptide derived by proteolysis of thyrotropin-releasing hormone. Although recent attention has focused on leptin and the ob gene, keep in mind that 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, “leptoς,” 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. But it was not until 1994 that the ob gene was identified, the gene responsible for obesity in the mouse. In the human, this gene is known as the Lep gene.

There are 4 recessive gene mutations known in mice, and one dominant (Ay/Ay). 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.

Leptin receptor

Genetic Rodent Models of Obesity

<table>
<thead>
<tr>
<th>Single Gene Mutations</th>
<th>Gene Product</th>
<th>Rodent Chromosome</th>
<th>Human Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mice: ob/+b</td>
<td>Leptin</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>ob/ob</td>
<td>Leptin receptor</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>db/db</td>
<td>Leptin receptor</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>fat/fat</td>
<td>Carboxypeptidase E</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>tub/tub</td>
<td>Phosphodiesterase</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Ay/Ay</td>
<td>Agouti protein</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Rats: fa/la</td>
<td>Leptin receptor</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Ob/ob and db/db mice were described over 30 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
The Leptin Receptor

The leptin receptor belongs to the cytokine receptor family. There are two major forms, a short form and a long form, OB-R_s and OB-R_l. 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 likely signaling receptor. The only place that the long form is expressed is in the hypothalamus, in the arcuate, ventromedial, paraventricular, and dorsomedial nuclei. However, a separate transport mechanism has also been demonstrated for leptin.

High levels of the short form leptin receptors in the choroid plexus suggest a transport role for the short form from blood into the cerebrospinal fluid to diffuse into the brain. However, a separate transport mechanism has also been demonstrated for leptin.

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 pathways. These signaling mechanisms are the subjects of intense molecular biology studies.

The db gene encodes the leptin receptor. The db mutation converts the long form to the short form. The db/db mouse has a single G to 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 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.

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. 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. 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.

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. Melanocortins binding to the brain receptor influence appetite. Knockout mice for the melanocortin receptor become obese, and these mice have high NPY expression. Agouti 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 adrenocorticotrophin (ACTH), yielding individuals with obesity, adrenal insufficiency, and red hair pigmentation.
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.

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.

In normal mice, leptin decreases NPY. 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. 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.

In the human, the ob gene, known as the Lep gene, does not appear to be acutely regulated. Feeding increases leptin levels, but slowly, correlating with insulin levels. Insulin provides negative feedback to the brain in a manner similar to leptin, including an effect on NPY. 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.

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, and therefore, lower leptin levels with fasting should lower CRH levels. 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. The specific food and energy response associated with CRH may also be mediated by urocortin, a CRH-related peptide. Urocortin is very potent in reducing food intake, perhaps activating the energy system without activating the overall CRH mechanism. These relationships are further confused by evidence that glucocorticoids stimulate leptin synthesis and secretion in humans.

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. 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. Higher levels in women suggest a greater resistance to leptin, correlating with a greater prevalence of obesity in women. The incidence of obesity in black women is almost twice that of white women. Obese postmenopausal black women have 20% lower leptin levels than white women. 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. 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. At least in the mouse, this resistance is present only in the periphery in that leptin administered in the brain still works.

Only a small percentage of humans are expected to have mutations in the leptin receptor and be refractory to leptin agonist treatment. Diet-induced obesity in mice responds to leptin treatment, and it is expected that the majority of overweight humans would also. However, in genetically normal animals, leptin treatment does not
always reduce weight by reducing food intake. A crucial question is what happens within physiologic range changes in leptin.

In obese individuals, the majority of leptin is unbound and presumably active, consistent with the resistance hypothesis. 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. 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, these 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.

Weight loss in obese and lean people produces the same response, decreased leptin and insulin. 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. When energy expenditure, food intake, and body weight are in balance, leptin is at a level consistent with a set point determined by this balance. With weight loss and loss of fat, leptin decreases, 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 levels change to restore the original status quo, making it difficult for overweight people to maintain weight loss. Thus, there is hope for leptin agonist treatment of these individuals.

Circulating levels of leptin are correlated with the percent body fat. 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.

A reduction of 10% in body weight is associated with a 53% reduction in serum leptin. 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.

Therefore, whenever a perturbation occurs, leptin levels change to restore the original status quo. Thus, these acute responses work against attempts to lose weight. A change in energy intake causes a change in leptin levels. 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. 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 have been rare, and mutations in the leptin receptor have also not been detected in obese people.

The first definitive demonstration of a congenital leptin deficiency in humans was reported in two severely obese Pakistani children who were cousins. 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 have been reported with an amino acid substitution that impairs intracellular transport of leptin. It is noteworthy that these individuals also displayed suppression of gonadal function.

An obese individual has been reported with a 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). However, a single gene mutation responsible for obesity will probably be discovered only in individuals such as these with extremely severe obesity. Nevertheless, 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.

It is expected that only a small percentage of obese humans will have mutations in the leptin receptor or the Lep gene.

Leptin and Reproduction

Several observations support a role for leptin in reproductive physiology.

1. Leptin administration accelerates the onset of puberty in rodents.
2. Leptin levels increase at puberty in boys.
3. Low leptin levels are present in athletes and in patients with anorexia and delayed puberty.
4. The ob/ob mouse undergoes normal sexual development, but remains prepubertal and never ovulates, but fertility is restored with leptin administration.

Leptin levels are greater in females than in males, and in premenopausal women compared with postmenopausal women. In girls, leptin levels are higher and decrease with increasing Tanner stages of puberty. 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.

The puzzle is why CRH is elevated in stress amenorrhea (especially that associated with weight loss) in contrast to 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. 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 hypercortisolism. 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. 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.

Although an initial study reported increased leptin levels in women with polycystic ovaries, the study has been criticized for not adjusting for body weight. At least 3 studies, controlling for weight, detected no differences in leptin levels comparing women with and without polycystic ovaries. 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.
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, age of 3, obesity in childhood predicts obesity in adulthood, and parental obesity doubles the risk of adult obesity in both obese and nonobese children.

Familial occurrence of obesity can be attributed in part to a genetically related reduced rate of energy expenditure.

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. 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. 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. Leptin, therefore, probably is involved in a multitude of important metabolic and developmental functions.

The leptin story has restored credibility to the critical weight hypothesis originally proposed by Rose Frisch in the 1970s. 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 has always been 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.

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, and the hypercortisolism is believed to be mediated through NPY signals in the hypothalamus. In stress-induced weight loss, CRH secretion is increased.

CRH directly inhibits hypothalamic 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. In regards 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) will depend on the degree of GnRH suppression.

A unifying hypothesis focuses on energy balance. 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. This is consistent with the finding that NPY stimulates GnRH secretion and amplifies gonadotropin response to GnRH in the rodent. From a teleological 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, the purpose of 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 apparent resistance to leptin action associated with excess body weight and fat would then reflect not resistance, but a lack of physiologic effect.

Although the leptin 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 or 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 may be 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. 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. 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. 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. 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 nonobese children.

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,

In these women.
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, a disorder which is not well understood. 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. 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. Increased body fat changes 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. 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. The simplest way to assess insulin resistance is to measure the ratio of fasting glucose to fasting insulin. A ratio lower than 4.5 is characteristic of insulin resistance. This method has limitations, the most notable being the variation due to assay precision and pulsatile secretion of insulin. It is more reliable to measure the insulin response to 1 g/kg glucose; the maximal response should not be greater than 150 μU/mL. However, if a patient is obese, one can assume the patient is insulin-resistant.

Genetics plays a greater role in the development of maturity onset diabetes than juvenile onset. 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. 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.

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.

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 adipocytokine 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 actively metabolizable. 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. In addition, women with central obesity have decreased cortisol levels, a finding that would be consistent with increased leptin levels. These metabolic changes improve with weight loss.

It is central body obesity that is associated with cardiovascular risk factors, including hypertension and adverse cholesterol-lipoprotein profiles. The waist:hip ratio is a variable strongly and inversely associated with the level of HDL, the fraction of HDL-cholesterol most consistently linked with protection from cardiovascular disease. 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. 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.

It is somewhat surprising that leptin expression and circulating levels are not influenced by android, central obesity, This indicates that leptin is regulated only by subcutaneous fat.

The waist:hip ratio is a means of estimating the degree of upper to lower body obesity; the ratio accurately predicts the amount of intra-abdominal fat (which is greater with android obesity). However, studies have demonstrated that the more easily determined circumference of the waist is a better predictor of central, android abdominal fat. A waist circumference greater than 100 cm (about 40 inches) in men and 90 cm (about 35 inches) in women is predictive of abnormal endocrinologic and metabolic function and is associated with an increased risk of cardiovascular disease.

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. 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.

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 pounds in the first month and 20–30 pounds in 4–5 months. To achieve a respectable rate of weight loss, intake must be 500–1000 calories below energy expenditure. 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 1200 calories per 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. There is no evidence supporting the contention that changing the relative proportions of carbohydrates, protein, and fat without reducing overall caloric intake will produce effective weight loss.

### Ideal Diet:

<table>
<thead>
<tr>
<th>Carbohydrates — 50%</th>
<th>Protein — 15–20%</th>
<th>Fat — &lt;30%</th>
</tr>
</thead>
</table>

The discouraging aspect is that to lose a pound of fat, the equivalent to a 3500-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. Restriction of calories remains the important principle, recognizing that reduction in fat intake is the most effective method of weight loss. Substituting one of the liquid formulas for meals has been successful in many individuals. Unbalanced formulations, however, have the same side effects as seen with total starvation (carbohydrate-deprived regimens). Adequate carbohydrate is necessary for utilization of amino acids. In addition, electrolyte problems have been encountered, and there is an initial diuretic phase that can lead to postural hypotension.

The protein-sparing modified fast is a ketogenic regimen providing approximately 800 calories per 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. 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 will have significant long-term weight reduction. 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 use of a dietian to record 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.

Anorectics are useful as short-term therapy to control hunger, especially at the beginning of a diet and at a plateau or relapse stage. Compared with amphetamines (which modulate both dopaminergic and noradrenergic neurotransmission), there is far less abuse associated with the catecholamine congeners that affect only noradrenergic neurotransmission. Fenfluramine and dexfenfluramine have no effects on either dopaminergic or noradrenergic neurotransmission, but they increase serotonin release and partially inhibit serotonin reuptake, thus depleting serotonin in the CNS by making it more available for metabolism. However, circulating levels of serotonin become elevated. All of these agents act on the central nervous system to depress appetite. 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 use for individuals who wish to lose a small amount of weight. Noradrenergic agents should not be used in individuals with cardiovascular disease. Dexfenfluramine is the d-enantiomer of fenfluramine, the active component of the racemic mixture of d,l-fenfluramine. It was reasoned that lower doses of a noradrenergic agent and a serotonin agent, when combined, would yield better results, and this proved to be true, but only in a small number of patients. The combination of phentermine and fenfluramine became popular, although weight is promptly regained when treatment is stopped. 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; primary pulmonary hypertension and cardiac valvular disease, either with the combination or with either agent (including dexfenfluramine) alone, and these products were withdrawn from the U.S. market. Symptoms associated with pulmonary hypertension include the onset of dyspnea, changes in exercise tolerance, angina, syncope, and lower extremity edema. Patients who develop valvular disease are usually symptomatic with dyspnea and evidence of congestive heart failure. Minor side effects of serotonergic drugs include diarrhea, polyuria, dry mouth, sleep disturbance, and somnolence. Patients who have used fenfluramine or dexfenfluramine should be examined, and those who have a heart murmur or who used the drugs for long durations or at high doses should be studied with echocardiography.

### Weight Loss Drugs

#### Noradrenergic Agents (affecting the catecholaminergic system):

- **Orlistat**: 120 mg tid
- **Phentermine**: 37.5 mg in morning
- **Mazindol**: 1–3 mg 1–3 times daily

#### Serotonergic Agents:

- **Sibutramine**: 10–15 mg daily
- **Dexfenfluramine**: 15 mg before meals

#### Noradrenergic and Serotonergic Agent:

- **Diethylpropion**: 25 mg before meals; 75 mg in morning in slow release form
- **Dexfenfluramine**: 120 mg tid

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 increase blood pressure. Orlistat (120 mg tid with meals) inhibits pancreatic lipase and increases fecal fat loss. Orlistat is associated with annoying gastrointestinal side effects (increased defecation of oily stools). There is some loss of the fat soluble vitamins, and a vitamin supplement should be taken at bedtime. 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. Although the serotoninergic agents produce neurotoxicity in animal brains, it is uncertain whether long-term depletion of brain serotonin has adverse neuropsychological effects (e.g., on mood and memory). Short-term studies indicate only a modest efficacy with variable responses. Most of the weight loss occurs in the first 6 months and is limited to 5–10 kg (11–22 lbs). 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. 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.

Over-the-counter products contain phenylpropanolamine as the active ingredient. This drug is a sympathomimetic derived from ephedrine and can act synergistically with caffeine to produce amphetamine-like reactions. It should be noted that phenylpropanolamine taken in combination with a dopamine agonist (e.g., bromocriptine) or a monoamine oxidase inhibitor can precipitate a hypertensive crisis. Appetite-suppressing drugs of all kinds are contraindicated in patients taking monoamine oxidase inhibitors and patients with glaucoma.

Surgical treatment and starvation should be reserved for patients who are morbidly obese. Both methods involve many potential problems and require close monitoring.

Controlled studies have not demonstrated the effectiveness of thyroid preparations or human chorionic gonadotropin. 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. Both weight loss and increased physical activity, through an unknown mechanism, lower the level of low density lipoprotein (LDL), and increase the level of high density lipoprotein (HDL). 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 hours. There is one study, however, that indicates a rebound increase in appetite 1–2 days after exercise. The optimal program includes, therefore, a daily period of exercise. A combination of diet and exercise is better than either alone, and those who exercise are more successful in maintaining weight loss. The best time for exercise is before meals or about 2 hours after eating. It is probably wise to take a day off at least once a week to give muscles and joints a rest.

Unfortunately one cannot burn up significant calories quickly; it takes 18 minutes of running to compensate for the average hamburger.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Calories per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleeping</td>
<td>90</td>
</tr>
<tr>
<td>Office work</td>
<td>240</td>
</tr>
<tr>
<td>Walking</td>
<td>240</td>
</tr>
<tr>
<td>Golf</td>
<td>300</td>
</tr>
<tr>
<td>Housework</td>
<td>300</td>
</tr>
<tr>
<td>Bicycling</td>
<td>360</td>
</tr>
<tr>
<td>Swimming</td>
<td>360</td>
</tr>
<tr>
<td>Tennis</td>
<td>480</td>
</tr>
<tr>
<td>Bowling</td>
<td>510</td>
</tr>
<tr>
<td>Cross country skiing</td>
<td>840</td>
</tr>
<tr>
<td>Running fast</td>
<td>960 (ca. 160/mile)</td>
</tr>
</tbody>
</table>

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 2700 calories to remain in equilibrium; spontaneously obese patients require only about 1300 calories. Others argue that virtually everyone can lose weight on a diet of 1000 calories per day in that the maintenance requirement for a sedentary adult is about 1.5 times the resting metabolic rate (about 1000–1500 calories per day). Nevertheless, an individual who has been overweight requires about 15% fewer calories to maintain weight than an individual who has never been obese. 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 appear doomed to frustration and despair unless the clinician can motivate them to increase physical activity. In all individuals, dieting is more effective when 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.

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.

Chapter References

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 (T\(_4\)) and triiodothyronine (T\(_3\)). 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 T\(_4\) yields T\(_3\), while removal of an iodine from the nonphenolic ring yields reverse T\(_3\) (RT\(_3\)) which is biologically inactive. In a normal adult, about one third of the T\(_3\) secreted each day is converted in peripheral tissues, largely liver and kidney, to T\(_2\), and about 40% is converted to the inactive, reverse T\(_3\). About 80% of the T\(_3\) generated is derived outside the thyroid gland, chiefly in the liver and kidney. T\(_3\) is 3–5 times more potent than T\(_4\), and virtually all the biologic activity of T\(_4\) can be attributed to the T\(_3\) generated from it. Although T\(_3\) is secreted at 20 times the rate of T\(_4\), it is T\(_3\) that is responsible for most if not all the thyroid action in the body. T\(_3\) is more potent than T\(_4\) because the nuclear thyroid receptor has a ten-fold greater affinity for T\(_3\) than T\(_4\). While T\(_3\) may have some intrinsic activity of its own, it serves mainly as a prohormone of T\(_3\). It is hard to think of a body process or function that doesn't 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 T\(_3\) levels in adults. A reciprocal relationship exists between T\(_3\) and RT\(_3\). Low T\(_3\) and elevated RT\(_3\) 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 T\(_2\) and RT\(_3\). During periods of stress, when a decrease in metabolic rate would conserve energy, the body produces more RT\(_3\) and less T\(_3\) and metabolism slows. Upon 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 T\(_4\) and thus allow T\(_3\) to have greater entry into cells. TBG is synthesized in the liver, and this synthesis is increased by estrogens. The nuclear receptor for thyroid hormone is a member of the super family that includes the steroid hormone receptors (Chapter 2). The thyroid hormone receptor exists in several forms, the products of 2 genes located on different chromosomes. The a receptor is on chromosome 17, and the b receptor is on chromosome 3. The nuclear T\(_3\) 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 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 T\(_3\) will produce a many-fold greater response in TSH. TSH-secreting cells are regulated by T\(_4\), but only after the T\(_4\) is converted to T\(_3\) 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 T\(_3\) for their intracellular T\(_3\),
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 combined oral 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. With aging, conversion of T₃ to T₄ 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.

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, 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 T₄/T₃)

Free thyroxine index is calculated from the TT₄ and the T₃ resin uptake measurements. This test too 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 where measurement will be useful will be discussed under the specific diseases and indicated on the algorithm.

Thyroid-Stimulating Hormone (TSH)

TSH is measured by highly sensitive assays utilizing monoclonal antibodies, usually in a technique that uses two antibodies, one directed at the a-subunit and one directed at the b-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. TSH is a very sensitive indicator of thyroid hormone action at the tissue level because it is dependent upon the pituitary exposure to T₄. In the absence of hypothalamic or pituitary disease, the sensitive TSH assays 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.

Radioactive Iodine Uptake

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 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 the TSH by a sensitive assay. 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 TSH will confirm the diagnosis of hyperthyroidism. If the T₄ is normal, the T₃ level is measured, since some patients with hyperthyroidism will have predominantly T₃ toxicosis. If the T₃ is normal, it implies that thyroxine secretion is autonomous from TSH, and this 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 the hypothyroidism is usually secondary to an autoimmune reaction, and when goiter formation is present, it is called Hashimoto’s thyroiditis. 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. It is uncertain whether hypothyroidism can be a cause of recurrent miscarriages, but an assessment of thyroid function is worthwhile in these patients.

Hypothyroidism increases with aging and is more common in women. Up to 45% of thyroid glands from women over age 60 show evidence of thyroiditis. The incidence of antithyroglobulin antibodies is 7.4% in women over 75 years, while 16.9% of women age 60 and 17.4% of women over age 75 have elevated TSH levels. 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 highly sensitive 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.
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 appear 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. 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.

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 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 will be 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 (usually 5–20 µU/L) and normal T₄ (called subclinical hypothyroidism). Many of these patients (but not all) will eventually become clinically hypothyroid with low T₄ concentrations. 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. An improvement in impaired cognitive function and emotional behavior has been documented with thyroxine treatment of subclinical hypothyroidism. In those patients who are asymptomatic, it is worth measuring antithyroid antibodies. A positive test identifies those who are likely to become clinically hypothyroid, at a rate of approximately 20% per year. With only very slight elevations of TSH (less than 10 µU/L), it is reasonable not to treat and to check thyroid function every 6 months to detect further deterioration. Patients with an abnormal cholesterol-lipoprotein profile can show improvement with thyroxine treatment, usually within one month of therapy.

**Treatment of Hypothyroidism**

Initial therapy is straightforward with synthetic thyroidine, 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. “Natural” thyroid preparations are not better, and in fact are potentially detrimental. Patients taking biological preparations should be switched to synthetic thyroxine. Because of a risk of coronary heart disease in older women, the initial dose should be 25–50 µg per 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. 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 6–8 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. A patient being treated with thyroid hormone should be evaluated once every year with the sensitive TSH assay. Thyroid hormone requirements tend to decrease with age. If the sensitive TSH assay is low, then the free T₄ should be measured to help adjust the thyroxine dose. The full response of TSH to changes in T₄ is relatively slow; a minimum of 6–8 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). Plummer’s disease is usually encountered in postmenopausal women who have had a long history of goiter. Twenty percent of hyperthyroid patients are over 60, and 25% of older women with hyperthyroidism present with an apathetic or atypical syndrome. Graves’ disease is characterized by the triad of hyperthyroidism, exophthalmos, and pretibial myxedema and is believed to be caused by autoantibodies that have TSH properties and, therefore, bind to and activate the TSH receptor. 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, heat intolerance, weight loss, sweating, palpitations, and diarrhea. These symptoms are associated with hyperthyroidism. Ophthalmopathy is rare in older patients. Hyperthyroidism in older women is sometimes described as “apathetic hyperthyroidism” because the clinical manifestations are different. 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 psychologic changes.

Psychologic 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 with a high T3 or a high T4 confirms the diagnosis. Hyperthyroidism caused by high levels of T3 is more common in older women. 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 cold 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 T3 and T4 levels, but subnormal concentrations of TSH. TSH levels can be suppressed to 0.1–0.5 µU/L by general illnesses and drugs such as glucocorticoids, dopamine, and anticonvulsants; however, this suppression does not extend below 0.1 µU/L. Values below 0.1 µU/L are regarded as nondetectable, and patients with overt hyperthyroidism usually have undetectable TSH. Subclinical hyperthyroidism is half as common in older people as clinical hyperthyroidism (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. If subclinical hyperthyroidism persists, it should be treated, especially in postmenopausal women, because of the cardiac complications and the loss of bone associated with excess thyroid hormone. TSH levels that are low but not undetectable (0.1–5.0 µU/L) need not be treated, but TSH measurement is warranted every 6 months. Progression to overt hyperthyroidism is uncommon.

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 which can exacerbate hyperthyroidism or be adversely affected by hyperthyroidism. Antithyroid drugs are usually administered first to achieve euthyroidism, then definitive therapy is accomplished by radioactive iodine treatment. Of course, it is important to make sure 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 will be methimazole because it has fewer adverse effects. The drug inhibits organification of iodide and decreases production of T3 and T4. 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 one 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 hours, orally, and the dose is titrated to maintain a heart rate of about 100 beats/minute. The drug may cause bronchospasm, worsening congestive heart failure, fatigue, and depression. Rarely inorganic iodine will be 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 radioidine administration for several months.

After the symptoms are controlled, and the patient is euthyroid, a radioactive iodine can be selected, the thioracil withheld temporarily, and definitive therapy accomplished. Patients with solitary nodules will be treated in the same fashion. Some patients with hot nodules in multinodular glands will 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 hyperthyroidism, the clinician needs to understand how thyroid hormone affects bones. Thyroid hormone excess alters bone integrity via direct effects on bone and gut absorption and indirectly through the effects of vitamin D, calcitriol, and parathyroid hormone. Thyroid hormone increases bone mineral resorption. In addition, total and ionized calcium increase in hyperthyroid women, leading to increases in serum phosphorous, 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 is increased bone resorption and a subsequent decrease in bone density — osteoporosis.

These effects become more clinically important in prolonged exposure to excessive thyroid hormone. Women who have had hyperthyroid experience postmenopausal fractures earlier than usual.

The major concern is that mild chronic excess thyroid hormone replacement, especially in postmenopausal women, might increase the risk of osteoporosis, and indeed this was subsequently documented. Bone density has been found to be reduced (9%) in premenopausal women receiving enough thyroxine to suppress TSH for 10 years or more. A meta-analysis of the literature on this subject concluded that premenopausal women treated for long durations did not suffer a clinically significant decrease in bone mass (probably because of the protective presence of estrogen); however, postmenopausal women lose an excess of bone if thyroid treatment results in TSH levels below the normal range. Thus, 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 the sensitive TSH assay to ensure that levothyroxine doses are “physiologic.” Some patients who require TSH suppressive doses of thyroid hormone, such as patients with nodules, goiters, and 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 bisphosphate 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.

Thyroid Nodules

The major concern with thyroid nodules is the potential for thyroid cancer. 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 from the age of 55. Mortality from thyroid cancer occurs predominantly in the middle-aged and 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 and 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. 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 of these findings, surgery is recommended for excision of the nodule. If none of these is present, proceed directly to fine needle aspiration biopsy or thyroid scan.

If a scan is chosen, hot nodules are evaluated independently. Cold nodules require a surgical diagnosis. If the patient prefers, one can treat with suppressive doses of levothyroxine and evaluate over time. Unfortunately, many of these thyroid nodules do not regress with thyroid treatment, but it is very reassuring if they do. Growth or lack of disappearance with thyroid suppression is an indication for fine needle aspiration biopsy. Thyroid ultrasound can be utilized to more accurately establish size for comparison over time.

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 the serum calcitonin as a very sensitive and specific tumor marker. Measuring the serum calcitonin is recommended for older patients with solitary nodules in order to achieve earlier diagnosis of medullary carcinoma.

**Fine-Needle Aspiration**

Fine-needle aspiration biopsy has a 83% sensitivity and 92% specificity in diagnosing thyroid malignancy. When “indeterminate” by biopsy, about one-third prove to be malignant at thyroideotomy. If the fine-needle aspiration biopsy reveals suspicious cells or is indeterminate, a subtotal thyroideotomy should be performed for diagnosis and treatment. If the aspiration biopsy is benign some would repeat the biopsy in one year to avoid false negatives. Most would provide thyroid hormone suppressive therapy (aiming for a TSH level below 0.1 µU/L) for one year with close observation for growth of the nodule. Growth or lack of disappearance indicates the need for biopsy or surgery. If the nodule shrinks during suppression, after one year a choice is made between no treatment or maintenance of the TSH level at the lower end of normal. In some cases, especially in older women, nodules can be followed with no therapy and little risk. If the nodule increases in size, suppressive therapy is recommended. One of the reasons many are treated with levothyroxine is because of the known growth-promoting effects of TSH in carcinoma of the thyroid and the hope that TSH suppression will inhibit the growth of early carcinoma.

The method for fine-needle aspiration biopsy requires no anesthetic. Using sterile technique and a 23-gauge needle on a 10 mL syringe, the nodule is fixed with two fingers of one hand and the nodule entered. Aspiration is performed with the syringe while several passes are made through a vertical distance of about 2 mm in the nodule. Suction is stopped when biopsy material becomes visible in the hub of the needle. The contents of the needle should be expelled onto a slide and fixed for pathology. Gentle pressure is applied over the nodule for 10 minutes. Occasionally, a patient will have some bleeding into the nodule or surrounding tissues, but it is usually self-limiting.

**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. However, a pregnant woman is euthyroid with normal levels of TSH, free T<sub>4</sub>, and free T<sub>3</sub>. 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. 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 compensated by an 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). It has been calculated that HCG contains approximately 1/4000th of the thyrotropic activity of human TSH. In conditions with very elevated HCG levels, the thyrotropic activity can be sufficient to produce hyperthyroidism, and this can even be encountered in normal pregnancy.

TBG levels reach a peak (twice nonpregnant levels) at about 15 weeks, which is maintained throughout the rest of pregnancy. Free T4 undergoes a similar change, but T3 increases more markedly. Because of the increase in TBG, free T4 and T3 levels actually decrease, although they remain within the normal range. There is an inverse relationship between maternal circulation levels of TSH and HCG. 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. These changes support a role for HCG stimulation of the maternal thyroid gland during early pregnancy. 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 severely limited in both directions. Indeed, the placenta is essentially impermeable to these substances no matter whether the fetus is euthyroid or hypothyroid. Slight transfer of T4 and T3 can occur, however, when maternal levels are very high or when fetal levels are substantially lower than the maternal levels.

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. 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). Specifically, HCG with reduced sialic acid content is increased in pregnant patients with hyperemesis and hyperthyroidism.

**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, the same time that the pituitary begins to synthesize TSH. 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.

The major thyroid hormone secreted by the fetus is T4. However, total T4 and free T4 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.
With delivery, the newborn moves from a state of relative T₄ deficiency to a state of T₃ thyrotoxicosis. Shortly after birth serum TSH concentrations increase rapidly to a peak at 30 minutes of age. They fall to baseline values by 48–72 hours. In response to this increase in TSH, total T₄ and free T₃ increase to peak values by 24–48 hours of age. T₃ levels increase even more, peaking by 24 hours 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 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 minutes. In some way cord cutting augments peripheral (largely liver) conversion of T₄ to T₃. The later increases in T₁ and T₃ (after 2 hours) 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. T₃ rises rapidly and exceeds maternal values at term.
3. T₂ levels rise, but concentrations are relatively low, similar to hypothyroid adults.
4. RT₁ levels exceed normal adult levels.
5. The fetal pattern of low T₄ and high RT₁ is similar to that seen with calorie malnutrition.
6. After delivery, TSH peaks at 30 minutes of age, followed by a T₁ peak at 24 hours and a T₄ peak at 24–48 hours. The T₃ increase is independent of the TSH change.
7. High RT₁ 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 one in 4000 live births, and newborn screening programs exist in most of the world. 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 treatment before 3 months of age is usually associated with normal mental development. Less than normal development can be a consequence of a delay in treatment or extremely low thyroid hormone production in the fetus.

There is a familial tendency for hypothyroidism, and if the diagnosis is made in the antepartum period, intra-amniotic injections of thyroxine can raise fetal levels of thyroid hormone. 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. However, fetal cord blood sampling is advocated for accurate diagnosis. Treatment of fetal hypothyroidism is important because there is a concern that 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, for a fetus with hypothyroidism.

Hyperthyroidism in Pregnancy

Untreated thyrotoxicosis in pregnancy is associated with a higher risk of preeclampsia, heart failure, intrauterine growth retardation, and stillbirth. Heart failure is a consequence of the demands of pregnancy superimposed upon the hyperdynamic cardiovascular state induced by the increased thyroid hormone. The most common cause of thyrotoxicosis in pregnancy is Graves’ disease. However, 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 upon 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.

The choice of treatment is between surgery and antithyroid drugs. However prior to surgery, the thyroid gland has to be controlled with medical therapy. Most women can be successfully treated with thioamide drugs. Propylthiouracil and methimazole are equally safe and effective for pregnant women; however, propylthiouracil is preferred in breastfeeding patients because it is less concentrated in breast milk.

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. 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. Nevertheless, pregnant women with thyrotoxicosis should be treated with as low a dose of antithyroid drugs as possible. With proper antithyroid drug treatment, very few, if any, deleterious effects are experienced by mother, fetus, or neonate. 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.

Maternal TSH-like autoantibodies 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 TSH-like antibodies to assess the fetal thyroid status. 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

This life-threatening augmentation of thyrotoxicosis 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 probably do not get pregnant. Patients with mild hypothyroidism probably never have a laboratory assessment for thyroid function during pregnancy and go undetected. Preeclampsia, intrauterine growth retardation, and fetal distress are frequent in women with significant hypothyroidism. There is also reason to believe that patients with hypothyroidism have an increased rate of spontaneous abortion. The mechanism may be impaired ability of important organs such as the endometrium and the corpus luteum. Women being treated for hypothyroidism require an increase (20–50%) in thyroxine during pregnancy. TSH should be monitored monthly in the first trimester and again in the postpartum period, and dosage should be adjusted to keep the TSH level in the normal range.

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%), usually 3–6 months after delivery, manifested by either hyperthyroidism or hypothyroidism, although commonly transient hypothyroidism is followed by hyperthyroidism. This condition is due to a destructive thyroiditis associated with thyroid microsomal autoantibodies. 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.

Most importantly, the symptoms in women who are often affected or depressed, 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. The symptoms of hypothyroidism in this condition are not responsive to antithyroid medication, and the patients are usually not treated or given beta-adrenergic blocking agents (e.g., propranolol in a dose sufficient to reduce the resting pulse to less than 100 per minute). Because spontaneous remission is common, patients who are treated with hypothyroidism should be reassessed one year after gradual withdrawal of thyroxine.

Patients who return to normal should undergo periodic laboratory surveillance of their thyroid status.

Chapter References

64. Kaplan MM, Monitoring thyroid hormone treatment during pregnancy, Thyroid 2:147, 1992.
As societies become more affluent, fertility decreases. This decrease is in response to the use of contraception and 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. In contrast, a modern urban woman will experience 420 menstrual cycles. Contemporary women undergo earlier menarche and start having sexual intercourse earlier 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 utilized.

More young women (under age 25) in the United States become pregnant than do their contemporaries in other Western countries. The teenage pregnancy rates in 5 northern European countries and Canada range from 13 to 53% of the U.S. rate. 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 78% of all pregnancies among American teenagers are unintended. American teenagers abort nearly half of their pregnancies, and this proportion is also seen in other countries. American women age 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. A decrease in unintended pregnancies (about 16%) and the abortion rate in the U.S. in the 1990s is due to better use of contraception; however, the rates are still relatively high. About half of all pregnancies in the U.S. are estimated to be unplanned, and more than half of these are aborted.

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 United Kingdom 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 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.

Contraception is not new, but its widespread development and application are new. The era of modern contraception dates from 1960 when oral contraception was first approved by the U.S. Food and Drug Administration, and intrauterine devices were re-introduced. 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.

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.

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 is defined as the number of failures per 100 woman-years of exposure. 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 1200 if the denominator consists of months or by 1300 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 one 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent of Women with Pregnancy</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No method</td>
<td>85.0%</td>
<td>85.0%</td>
</tr>
<tr>
<td>Combination Pill</td>
<td>0.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Progesterin only</td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>IUDs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 rates) as noted in the table of failure rates during the first year of use. The lowest expected failure rates are determined in clinical trials, where the combination of highly motivated subjects and frequent support from the study personnel yields the best results. It should be noted that slightly more than half of the unintended pregnancies in the U.S. are due to contraceptive failures.

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 and 1995. The samples are very large, and, therefore, the estimates are very accurate.

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, and then decreased in 1995, especially among Hispanic and black Americans. Approximately 10.7 million American women used oral contraceptives in 1988, and 10.4 million in 1995. Among never married women, oral contraception has been the leading method of birth control, but from 1988 to 1995, oral contraceptive use decreased in women younger than 25 and rose among women aged 30–44. A part of the decrease in oral contraceptive use is due to the new availability and use of implant (about 0.5 million women in 1995) and injectable methods (about 1.0 million women in 1995). However, the greater impact is due to an increase in condom use, especially by never married and formerly married women, women younger than 25, black women, and Hispanic women; indeed, the recent increase in overall contraceptive use is due to the increase in condom use which rose from 5.1 million women aged 15–44 in 1988 (15%) to 7.9 million in 1995 (20%). This increase occurred in all race and ethnic groups. These changes reflect the concern regarding sexually transmitted diseases, including human immunodeficiency (HIV). About one-third of condom users in 1995 were using more than one method, especially younger and never married women!

In 1982, 56% of U.S. women, 15–44 years of age, were using contraception, and this increased in 1995 to 64% (about 39 million women). In 1995, contraceptive sterilization (male and female) was utilized by 36% (15 million) of these women (the next leading method was oral contraception, 27%) The number of reproductive age women using the intrauterine device (IUD) decreased by two-thirds from 1992 to 1988, and further decreased in 1995, from 7.1% to 2% to 0.8%, respectively. IUD use is concentrated in the U.S. in married women over age 35. In 1982 more than 2 million women used the IUD and a similar number used the diaphragm, but in 1995 only 0.3 million women used the IUD and 0.7 million women (1.9%) used the diaphragm. The decrease in IUD use since 1982 has been especially marked in Hispanic and black women.
The oral contraceptive is the most popular method 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. This warrants a renewed effort to extend the use of these methods.

In 1995, 95% of those who were at risk of getting pregnant were using some method of contraception, whereas 36% of women of reproductive age were not using a method of contraception for the following reasons:

- 11%—no sexual experience,
- 9%—pregnant or trying to get pregnant,
- 6%—not sexually active,
- 5%—sterilized for medical reasons,
- 5%—at risk for an unintended pregnancy (this percentage steadily decreased over the last decade).

U.S. couples have made up for the lack of contraceptive choices 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 20% of couples in 1995.

The pattern of contraceptive use in Canada in 1995 was similar to that of the U.S., with a slightly higher percentage of oral contraceptive use (38.2% of women 15–44 years of age) and a slightly lower utilization of sterilization (32.5%). Canada, too, has seen an increase in condom use and a decrease in use of the IUD. In France, 36.3% of reproductive aged women used oral contraceptives (in 1994), 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. 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 in 1994 — male sterilization is virtually nonexistent).

The Impact of the Worldwide Use of Contraception

The world population is expected to stabilize at between 11 and 12 billion around 2150, with a fertility rate of 2.1 children per woman. Approximately 95% of the growth will occur in developing countries, so that by 2100, 13% of the population will live in developed countries, a decrease from the current 25%.

Throughout the world, 45% of married women of reproductive age practice contraception. However, there is significant variation from area to area; e.g., 69% in east
Throughout the world, 45% of married women of reproductive age practice contraception. However, there is significant variation from area to area; e.g., 69% in east Asia but only 11% in Africa. Female sterilization and the IUD are most popular in developing countries, whereas oral contraceptives and condoms are most popular in developed countries. 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 in a developing country than in a developed country. The health risks associated with pregnancy and childbirth in the developing world are far greater than risks secondary to the use of modern 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. It is not enough to simply limit fertility; contraception is only one component of reproductive health.

The Impact of Use and Non-use

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. Both safe and unsafe abortions can be minimized by maximizing contraceptive services. However, the need for safe abortion services will persist. Contraceptive failures account for about half of the 1.33 million annual induced abortions in the U.S.

In the U.S. in the late 1980s, $1 spent on public funding for family planning saved an average of $4.40 spent on medical, welfare, and nutritional services. States with higher family planning expenditures have fewer induced abortions, low-birthweight newborns, and premature births. 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 impacts largely on poor women, increasing the number of unintended births and abortions. In California, in the 1980s, there was an average of $7.70 saved for every $1 spent to provide contraceptive services.

There is a gap between the low levels of unintentional pregnancy that can be achieved and the actual levels being obtained, most of which are in couples using reversible contraception. A major thrust, in addition to providing services, must include education and counseling of couples about effective contraception.

STDs and Contraception

The interaction between clinician and patient for the purpose of contraception provides an opportunity to control sexually transmitted diseases (STDs). The World Health Organization estimates that in the year 2000, over 13 million women will be infected with the human immunodeficiency virus (HIV). 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 STD 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 STDs 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 STDs.

The Future

From 1970 to 1986, the number of births in women over 30 quadrupled; since 1990, the fertility rate among women over 30 has remained relatively stable. As more and more couples defer pregnancy until later in life, the use of sterilization under age 35 will decline, and the need for reversible contraception will increase. Between 1988 and 1995, oral contraceptive use decreased in women younger than 25 and increased in women aged 30–44. These numbers changed because clinicians and patients have come to understand and accept that low-dose oral contraception is safe for healthy, nonsmoking older women.

The need for reversible contraception in women over the age of 30 is growing, not diminishing. 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 not reach their 45th birthday until around 2010. For approximately a 20-year period, therefore, there will be an unprecedented number of women in the later childbearing years. The number of women aged 35–49 increased 61% between 1982 and 1995. The proportion of births accounted for by this group of women will increase by about 72%, from 5% in 1982 to 8.6% in 2000. This group of women is not only increasing in number, but it is changing its fertility pattern.

The deferment of marriage is a significant change in our society. In 1960, 28% of women aged 20–24 were single; in 1985, 58.5%. In 1960, 10% of women aged 25–29 were single; in 1985, 26%. But only 16% of the decline in the total fertility rate is accounted for by the increase in the average age at first marriage. Eighty-three percent of the decline in total fertility rate is accounted for by changes in marital fertility rates. In other words, postponement of marriage 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 will be seeing more and more older women who will need reversible contraception. In short, there will continue to be longer duration of use in younger women and greater use in older women, a pattern that began in 1990.

The first column of the table shows the projected births to women aged 15–44, as indicated in the U.S. Department of Commerce, Bureau of the Census, Current Population Reports. The second column shows the number of women aged 15–44 who would have had an induced abortion if they had not used contraception. The third column shows the number of women who would have had an induced abortion if they had not used contraception and if they had used contraception more effectively. The fourth column shows the number of women who would have had an induced abortion if they had not used contraception and if they had used contraception more effectively and if the abortion rate had been reduced by 50%. The last column shows the number of women who would have had an induced abortion if they had not used contraception and if they had used contraception more effectively and if the abortion rate had been reduced by 50% and if the contraceptive failure rate had been reduced by 50%.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–24</td>
<td>19.5 mill.</td>
<td>17.4 mill.</td>
<td>16.7 mill.</td>
<td>17.7 mill.</td>
</tr>
<tr>
<td>25–29</td>
<td>10.9</td>
<td>10.6</td>
<td>9.3</td>
<td>8.6</td>
</tr>
<tr>
<td>30–34</td>
<td>10.0</td>
<td>11.0</td>
<td>10.8</td>
<td>9.4</td>
</tr>
<tr>
<td>35–44</td>
<td>16.2</td>
<td>19.1</td>
<td>21.1</td>
<td>21.9</td>
</tr>
<tr>
<td>Total</td>
<td>56.6</td>
<td>58.1</td>
<td>57.9</td>
<td>57.6</td>
</tr>
</tbody>
</table>

One solution to the problem of a restricted number of contraceptive choices for American women is to develop new methods. There are many obstacles to the development of new methods, including the attitudes of the American public (besides America's traditionally conservative, religion-oriented views toward sex and family, polarization is produced by responses evoked by specific issues such as sterilization and induced abortion), the funding available for research, the time and cost required to meet federal regulations, and the problems of product liability. However, experts in the field remain optimistic, citing the many needs and opportunities for improvements and new developments.

Fortunately, clinicians and patients have recognized that low-dose oral 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 over age 40. However, as the above statistics indicate, its use is still not sufficient to meet the need. Besides fulfilling a need, this population of women has a series of benefits to be derived from oral 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 utilization of the IUD. 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 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...
Laparoscopy and minilaparotomy have led to a profound change in the convenience and cost of sterilization operations for women. In 1970, the average woman in the U.S. These methods have allowed women to undergo sterilization operations at times other than immediately after childbirth or during major surgery.

Since 1975, minilaparotomy, a technique popular in the less developed world, has been increasingly performed in the U.S. This same trend has occurred in Great Britain, where by age 40, over 20% of men and women have had a sterilization procedure. In Spain and Italy, sterilization rates are very low, but the use of oral contraceptives and the IUD is very high.

The growing need for reversible contraception would also be served by increased utilization of the IUD. 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 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. Unfortunately, a significant percentage of women still do not know that there are many health benefits with the use of oral 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. At the turn of the 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. But 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, nearly one million Americans each year have undergone a sterilization operation, and recently, more women than men. By 1995, 38% of reproductive aged women relied on contraceptive sterilization: 27% had tubal occlusion (11 million women), and 11% depended on their partners' vasectomies (4 million men). This same trend has occurred in Great Britain, where by age 40, over 20% of men and women have had a sterilization procedure. In Spain and Italy, sterilization rates are very low, but the use of oral contraceptives and the IUD is very high.

**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. He also proposed a technique for sterilization, which he later described so precisely that he must 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. 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 over 40 and to have had 8 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.

By the end of the 1960s, sterilization was a popular procedure.

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 due to 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. 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.

Laparoscopy and minilaparotomy have led to a profound change in the convenience and cost of sterilization operations for women. In 1970, the average woman...
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. 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 childbirth (about 10 per 100,000 births in the U.S.). 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. When patients consider sterilization, we can assure them that vasectomy has not been demonstrated to have any harmful effects on men’s health. 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. After 6 weeks or 15 ejaculations, essentially all men are sterile.

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent of Women with Pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest Expected</td>
</tr>
<tr>
<td>Female sterilization</td>
<td>0.2%</td>
</tr>
<tr>
<td>Male sterilization</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Besides 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. 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. 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.

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. Bipolar tubal coagulation is more likely to result in ectopic pregnancy than is mechanical occlusion. 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 2 or more years after sterilization, rather than immediately after. In the first year after sterilization, about 6% of pregnancies will be ectopic, but the majority of pregnancies that occur 2–3 years after occlusion will be ectopic. 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.

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%. 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

Evidence from the Nurses’ Health Study indicated that tubal sterilization was associated with a 67% reduced risk of ovarian cancer. 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. In addition, case-control data have consistently supported this finding. The mechanism for such an effect is not clear, but this information is worth sharing with patients.

### Female Sterilization Techniques

- **Tubal Ligation**
  - Minilaparotomy
  - Laparoscopy
  - Transvaginal sterilization

- **Vasectomy**
  - In-office outpatient procedure
  - Hospital outpatient procedure
  - Same day surgery

- **Implant**
  - Implantation of a small, non-implanted device into the woman's arm

- **Intrauterine Device (IUD)**
  - Inserted into the uterus

- **Hormonal Methods**
  - Contraceptive pills
  - Contraceptive patches
  - Contraceptive rings
  - Contraceptive implants

- **Barrier Methods**
  - Condoms
  - Diaphragms

- **Cultural Considerations**
  - Respect patient preferences and choices
  - Provide comprehensive counseling about the method

- **Monitoring and Follow-up**
  - Regular check-ups to ensure method effectiveness
  - Counseling for patients considering reversing sterilization

The choice of contraceptive method should be carefully discussed with patients to ensure they understand the benefits, risks, and effectiveness of each option.
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 hours. 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, or 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 women experience a failure within 10 years. As much as one-third of these failures are ectopic pregnancies. The higher failure rates with silastic rings, clips, and bipolar coagulation reflect the greater degree of skill required for these methods. The methods with a greater percentage of failures occurring many years after the procedure are unipolar coagulation, bipolar coagulation, and the silastic ring. 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, women seeking sterilization should consider the use of the copper IUD or implants, reversible methods that offer very low failure rates.

### Female Tubal Sterilization Methods, 10-Year Cumulative Failure Rates

<table>
<thead>
<tr>
<th>Method</th>
<th>Failure Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unipolar coagulation</td>
<td>0.75%</td>
</tr>
<tr>
<td>Postpartum tubal excision</td>
<td>0.75</td>
</tr>
<tr>
<td>Silastic (Falope or Yoon) ring</td>
<td>1.77</td>
</tr>
<tr>
<td>Interval tubal excision</td>
<td>2.01</td>
</tr>
<tr>
<td>Bipolar coagulation</td>
<td>2.48</td>
</tr>
<tr>
<td>Hulka-Clemens clip</td>
<td>3.65</td>
</tr>
</tbody>
</table>

### 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. In order to complete the circuit, however, these electrons must spread through the body and be returned to the generator via a return electrode 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.

With the unipolar method, if tissue resistance is high and the electrical pressure (voltage) relatively low, current may cease to flow or may search out alternate pathways with lower resistance. When the voltage is increased, the electrons have more "push" to find another pathway, therefore, the surgeon must use the lowest possible voltage necessary to completely coagulate. The return electrode (the ground plate) must be in good contact with the patient.

Unipolar electrocautery can create a unique electrical "capacitance" problem when an operating laparoscope is used with unipolar forceps. 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 have occurred with laparoscopic sterilization. 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. Should bowel or other structures touch a laparoscope, which is insulated from the abdominal incision (for example, by a fiberglass 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-like fiberglass. 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 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 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 utilized. The transaction 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 pathological 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. In addition, the desiccated tissue can adhere to the bipolar forceps, making it difficult to remove from the surface of the tube.

The bipolar method can be used with either a single-incision operating laparoscope or with dual incision instruments. The forceps are, however, more delicate than unipolar equipment and must be kept meticulously clean. Damage to the instruments can alter the ability to coagulate, and inadequate or incomplete electrocoagulation is the main cause of failure.

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 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. In addition, case-control data have consistently supported this finding. The mechanism for such an effect is not clear, but this information is worth sharing with patients.
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 should be 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 one-year pregnancy rates of 2 per 1000 women, but the highest 10-year cumulative failure rate. 44, 59 Complications unique to spring clip sterilization result from mechanical difficulties. Should the clip be 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 is destroyed. Failure rates one year after the procedure with the newest model approximate 1 per 1000 women. 49 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.

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 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 ring applicator consists of two concentric cylinders. Within the inner cylinder is a forceps for grasping, elevating, and retracting a segment of the tube. The silastic ring is stretched around the exposed end of the inner cylinder by means of a special ring loader and ring guide. The outer cylinder moves the ring from the inner cylinder on to the tube, a loop of which is held within the inner cylinder by the forceps.

As with application of clips, the ring should be placed at the junction of the proximal and middle third of each fallopian tube. Once the tube is grasped, it is gently withdrawn into the inner cylinder by slowly squeezing the pistol-like handle of the applicator. A final strong pull is needed to slide the ring from the inner applicator.
Young women in unstable relationships need special attention in counseling, and both partners should participate in the counseling. Furthermore, for many couples years.

Reversibility

More studies with careful attention to the type of tubal occlusion procedure will be necessary. The best answer for now is that some women experience menstrual with longer (4 years) follow-up, while no changes have been noted with the use of rings or clips.

It is possible that extensive electrocoagulation of the fallopian tubes can change ovarian steroid production. Perhaps this is why menstrual changes were detected sterilization was concentrated in women who were treated for gynecologic disorders before tubal sterilization.

Menstrual Function

The effects of tubal sterilization on menstrual function are less clear, and, therefore, more difficult to explain. The first well-controlled studies of this issue demonstrated no change in menstrual patterns, volume, or pain. Subsequently these same authors reported an increase in dysmenorrhea and changes in menstrual bleeding. However, these authors failed to agree in their findings (a change found by one group was not confirmed by the other). Adding to the complexity, the incidence of hysterectomy for bleeding disorders in women after tubal sterilization was reported to be increased by some, but not by others. In a large cohort of women in a group health plan, hospitalization for menstrual disorders after unipolar cautery cannot be documented. Still another long-term follow-up study (3–4.5 years) failed to document any significant changes in menstrual cycles. This inconsistency can reflect differences in sterilization techniques, as well as the fact that a surgical solution is more likely to be chosen if continuing fertility is no longer an issue.

More studies with careful attention to the type of tubal occlusion procedure will be necessary. The best answer for now is that some women experience menstrual changes, but most do not.

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. Not all couples are pleased following sterilization; in one series, 2% of U.S. women expressed regret one year later and 2.7% after two years. At the two-year mark, the main factors associated with regret were age less than 30 and sterilization at the convenient time of a cesarean section. In Europe where tubal sterilization is less common, the most important risk factor for regret was an unhappy marriage. A change in marital status is undoubtedly an important reason for a desire to reverse sterilization.

Young women in unstable relationships need special attention in counseling, and both partners should participate in the 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. Microsurgery for tubal anastomosis is associated with excellent results if only a small segment of the tube has been damaged. Pregnancy rates correlate with the
Young women in unstable relationships need special attention in counseling, and both partners should participate in the 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. 5 2 5

Microsurgery for tubal anastomosis 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. 5 2 2 About 2 per 1000 sterilized women will eventually undergo tubal anastomosis. 5 3

Male Sterilization: Vasectomy

Vasectomy is safer, easier, less expensive, and has a lower failure rate than female sterilization. 5 4 The operation is almost always performed under local anesthesia, usually by a urologist in a private office. 5 5 Surgeons who do more than 10 operations yearly have lower complication rates. 5 5

Hematoma 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 cardiovascular disease. 5 5 12 Adverse psychological and sexual effects have not been reported. Since the other components 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 11 in the United States. An increased risk of prostate cancer after vasectomy has been reported in several cohort and case-control studies. 5 6 5 7 5 11 5 12 However, there is disagreement because other studies could not support an association between prostate or testicular cancer risk and vasectomy. 5 3 5 4 5 5 5 7 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. 5 7 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. 5 11 5 12 A meta-analysis of the literature concluded there is no increased risk of prostate cancer in men who have undergone vasectomy. 5 12 Observational studies cannot totally avoid potential biases, and the disagreement in regards to 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.

Vasectomy reversal is associated with pregnancy rates as high as 70–80%. 5 11 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. 5 12

Induced Abortion

Contraception is more effective and convenient than ever, but our modern methods are far from perfect. Even the most conscientious couples can experience contraceptive failure. In the past, failure of contraception meant abortion or, 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" decisions 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 number of abortions performed in the United States has been decreasing since a peak was reached in 1990, totaling 1.33 million in 1993 and 1.21 million in 1995. 5 10 5 11 This is partly because the number of pregnancies in the U.S. has been decreasing and the proportion of reproductive age women under age 30 is also decreasing. 5 10 About one-third of pregnancies not ending in miscarriage or stillbirth are terminated by abortion. The proportion of abortions performed in hospitals has steadily declined, reaching 7% in 1992. The proportion handled by specialized abortion clinics has increased, while the percentage of abortions performed by physicians in their own offices has remained low, about 3-5% of all abortions. More than 50% of abortions are obtained by women younger than 25 (about 20% under the age of 20), with the rate peaking at ages 18–19, and about 80% are unmarried. 5 10 5 11

A concern in the U.S. is the fact that the decline in the annual number of induced abortions has been greater than the decline in the annual number of births. 5 10, 5 11 Possible explanations include a reduction in unintended pregnancies (better contraception) and a greater willingness to experience a pregnancy; however, another possibility is reduced access to abortion services.

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 1993, 20% of women who obtained legal abortions were adolescents. 5 12 In addition, from ages 20–34, American women have the highest proportion of pregnancies aborted compared with other countries, indicating an unappreciated, but real problem, of unintended pregnancy occurring beyond the teenage years. The lack of perfect contraception and imperfect use of contraception will keep abortion with us.

Public health authorities have demonstrated that the legalization of abortion reduced maternal morbidity and mortality more than any single development since the 19th century. The number of pregnancies not ending in miscarriage or stillbirth are terminated by abortion. The proportion of abortions performed in hospitals has steadily declined, reaching 7% in 1992. The proportion handled by specialized abortion clinics has increased, while the percentage of abortions performed by physicians in their own offices has remained low, about 3-5% of all abortions. More than 50% of abortions are obtained by women younger than 25 (about 20% under the age of 20), with the rate peaking at ages 18–19, and about 80% are unmarried. 5 10 5 11

A concern in the U.S. is the fact that the decline in the annual number of induced abortions has been greater than the decline in the annual number of births. 5 10, 5 11 Possible explanations include a reduction in unintended pregnancies (better contraception) and a greater willingness to experience a pregnancy; however, another possibility is reduced access to abortion services.

Major and minor complications in a series of 170,000 first-trimester abortion patients were as follows: 5 12

**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
- Resuspension same day — 220.8
- Resuspension 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 over 150 studies. 5 12 There is no evidence for any adverse consequences of vacuum aspiration abortion for subsequent fertility. 5 13, 5 17 pregnancies, 5 13, 5 17 or increased risk for ectopic pregnancy. 5 13, 5 17 It is not yet clear if second-trimester abortions or several first-trimester abortions can affect the outcome of later pregnancies. However, multiple induced abortions do not increase the risk of a subsequent ectopic pregnancy. 5 13 The long-term effects of second-trimester abortion may depend on the method used. 5 13 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. 5 13
can be exploited for many uses, including contraception, therapy of endometriosis, induction of labor, treatment of Cushing's syndrome, and, potentially, treatment of ectopic pregnancy. Mifepristone is most noted for its abortifacient activity and the political controversy surrounding it. However, the combination of its agonistic and antagonistic actions has led to its use in various therapeutic applications.

Developed, but only mifepristone has undergone extensive abortion trials. It seems unlikely that mifepristone could have serious adverse effects, and there have been no reports of severe complications. Misoprostol (administered vaginally) achieved a 95% complete abortion rate in women at 9–13 weeks' gestation. 

Aspiration 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 it is performed. Counseling has played a critical role in the development of efficient and acceptable abortion services. Counseling abortion patients includes providing information about the procedure, obtaining informed consent, provision of emotional support for the patient and her family before, during, and after the operation, and providing information about contraception. Referral opportunities should be provided for pregnant 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.

Methods for First-Trimester Abortions

The most widely used technique for first-trimester abortions is vacuum curettage (99% of legal induced abortions in 1995). The procedure is performed using local anesthesia (a paracervical block). Cervical dilation is accomplished with tapered Pratt dilators. Some physicians 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).

Mifepristone (RU486), the progesterone antagonist, produces prooperative cervical dilation equally effectively, and the ease of its single oral dosage makes it a more attractive choice. After the procedure, the patient is observed for 1–2 hours before returning home.

Aspiration 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 it may be subject to loss of privacy or harassment. Nonsurgical methods might make abortion available to more women and improve the circumstances under which pregnancies are terminated. Two such methods have undergone clinical testing. The progesterone antagonist mifepristone (RU486) and the antimitabolite methotrexate have both been demonstrated to induce abortion early in pregnancy when combined with a prostaglandin.

France and China were the first countries to approve the marketing of the medical abortifacient mifepristone (now available in Great Britain and Sweden as well), a synthetic relative of the gestational agents in oral contraceptives. Mifepristone acts primarily, but not totally, as an antiprogesterational 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 progesterin antagonist is due to its ability to achieve this conformational change and the transcription activation functions on the progesterone receptor. The dimethyl (dimethylaminophenyl) side chain at carbon 11 is the principal factor in its abortifacient 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 600 mg oral dose of mifepristone is followed a day later by the administration of a prostaglandin analogue. Several analogues have been used, but the most widely available and best tolerated is mifepristone, 800 mg administered vaginally. The combination allows a reduction in dosage of both agents. When administered in the first 8 weeks of pregnancy, this medical termination carries success and complication rates similar to that achieved with vacuum curettage.

Mifepristone (RU486), a stable, orally active synthetic analogue of prostaglandin E1, available commercially for the treatment of peptic ulcer. Combined with misoprostol, it provides an effective, simple, inexpensive, completely oral or vaginal method. In the large U.S. trial of 600 mg mifepristone followed by 400 mg misoprostol orally, there was a 1% failure rate under 7 weeks of pregnancy and 9% from 8 weeks to 9 weeks. Based on worldwide experience, the regimen with the least side effects and cost, but equally good efficacy, is a combination of a lower dose of oral mifepristone (200 mg), 36–48 hrs later by the vaginal administration of 800 mg misoprostol.

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 withdrawal of support from the corpus luteum — the more dependent the pregnancy is on progesterone from the corpus luteum, the more likely 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. Other progesterone antagonists have been developed, but only mifepristone has undergone extensive abortion trials. It seems unlikely that mifepristone could have serious adverse effects, and there have been none reported.

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.
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.

Lack of progesterone antagonists in the U.S. prompted use of methotrexate as an abortifacient in the same dose used to treat ectopic pregnancy, 50 mg intramuscularly per square meter of body surface area. Later, a single 75 mg intramuscular dose was demonstrated to be as effective. Methotrexate has also been administered orally, in doses of 25 or 50 mg. As with mifepristone, a prostaglandin is added to promote expulsion of the uterine contents, and again vaginal misoprostol is the most useful analogue. The first trials demonstrated that if the prostaglandin (800 mg misoprostol vaginally) was given a week after the injection of methotrexate, this method could be almost as effective as mifepristone. Like mifepristone, efficacy diminishes with advancing gestation beyond 7 weeks (since the last menstrual period). 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. There are no direct comparison studies of methotrexate and mifepristone.

Another new approach uses the combination of tamoxifen and misoprostol. The administration of tamoxifen (20 mg daily for 4 days) followed by misoprostol (800 mg vaginally, with a second dose if necessary 24 hours later) was associated with a 92% rate of complete abortion in 100 women with pregnancies less than 9 weeks gestational age. The use of prostaglandin alone has also been pursued. Relatively high success rates have been reported with multiple dosing, but the most effective regimen and the best method of administration remain to be determined. The administration of 800 mg misoprostol daily for 3 days has been reported to be very effective late in the first-trimester (10–12 weeks).

One word of caution regarding misoprostol, the synthetic prostaglandin E₁ 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 echocardiocerebral and arthrogryposis. 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.

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 may 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%. Patients with retained products of conception occasionally present several weeks after an abortion, but most report excessive bleeding within one 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. Some studies indicate that prophylactic antibiotics reduce the risk of postabortal infection. 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, while still others believe that all abortion patients would benefit from prophylactic antibiotics. A meta-analysis of antibiotics at the time of induced abortion unequivocally concluded that prophylactic antibiotics should be routinely used without exceptions. 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 abortion and 200 mg 30 minutes afterward, is the most convenient and comprehensive regimen.

Tetracycline, 500 mg once before and once after the operation, is also acceptable. Metronidazole, 400 mg an hour before and 4–8 hours afterward, has been tested and is effective treatment for patients with bacterial vaginosis detected at the time of abortion.

Patients who present with uterine tenderness, fever, and bleeding require uterine aspiration as well as antibiotic treatment. Patients who have fevers above 38°C (101°F) and signs of peritoneal inflammation, as well as uterine tenderness, require hospitalization and intravenous antibiotics active against anaerobes, gonorrhea, and chlamydia. Outpatient treatment with doxycycline, 100 mg bid 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 examination of chorionic gonadotropin and vaginal ultrasonography should be used for accurate diagnosis and management.

Cervical Stenosis

Patients who experience amenorrhea or hyomenorrhoea 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 also 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 mg Rh immunoglobulin to all Rh-negative, Du-negative women undergoing early abortion. The standard dose is administered for second-trimester abortion.

Abortion in the Second-Trimester

Second-trimester abortions can be accomplished surgically or medically. The surgical procedure is termed dilatation 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. The D & E procedure is safer and less expensive than the medical methods and is better tolerated (and thus
Abortion in the Second-Trimester

Second-trimester abortions can be accomplished surgically or medically. The surgical procedure is termed dilation and evacuation (D & E). Several approaches have been used for the medical treatment of pregnancy. These include the vaginal, intramural, or intravascular administration of prostaglandins and the intra-uterine injection of hypertonic saline or urea. The D & E procedure is safer and less expensive than the medical methods and is better tolerated (and thus preferred) by patients. 34-36

In 1994, only 4% of induced abortions occurred at 16-20 weeks of gestation and 1% at 21 weeks or later. 37 In 1994, 94% of second-trimester abortions were performed by D & E.

The training, experience, and skills of the surgeons 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 the capacity of the equipment. Uterine perforation, infection, inflammation, anoxic fluid embolism, and hemorrhagic reactions are increased as gestational age increases.

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 overdose. 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 are compounded by the pre-emptive intervention of the surgeon with uterine evacuation techniques, the patient may undergo unnecessarily extensive surgery, such as hysterectomy, with all the risks inherent in emergency procedures.

Preoperative cervical dilatation with osmic 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. Some patients are not good candidates for surgical procedures of any kind under local anesthesia. 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.

In the United Kingdom, prostaglandin analogues are favored for a noninvasive method of second-trimester abortion. Attempts to use the prostaglandin antagonist, meflopristone, (a single oral 200 mg dose of mifepristone administered 36 hours before prostaglandin treatment) and an E prostaglandin analogue (misoprostol) administered orally or vaginally is highly effective, and the combination allows a lesser dose of both agents, which results in fewer side effects. In addition, this combination does not require the use of cervical laminaria for dilatation.

Chapter References

25. Peterson HB, Xia Z, Wingo J,aggregate the death rates from local to general anesthesia, the patient is at much greater risk of anesthetic complications. Finally, if complications caused by advanced gestational age are compounded by the pre-emptive intervention of the surgeon with uterine evacuation techniques, the patient may undergo unnecessarily extensive surgery, such as hysterectomy, with all the risks inherent in emergency procedures.

Preoperative cervical dilatation with osmic 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. Some patients are not good candidates for surgical procedures of any kind under local anesthesia. 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.

In the United Kingdom, prostaglandin analogues are favored for a noninvasive method of second-trimester abortion. Attempts to use the prostaglandin antagonist, meflopristone, (a single oral 200 mg dose of mifepristone administered 36 hours before prostaglandin treatment) and an E prostaglandin analogue (misoprostol) administered orally or vaginally is highly effective, and the combination allows a lesser dose of both agents, which results in fewer side effects. In addition, this combination does not require the use of cervical laminaria for dilatation.


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 oral contraception with synthetic sex steroids that is recent.
It wasn't until the early 1900s, that inhibition of ovulation was observed to be linked to pregnancy and the corpus luteum. Ludwig Haberlandt, professor of physiology at the University of Innsbruck, Austria, was the first to demonstrate that ovarian extracts given orally could prevent fertility (in mice). In the 1920s, Haberlandt and a Viennese gynecologist, Otfried Otto Fellner, were administering steroid extracts to a variety of animals and reporting the inhibition of fertility. By 1931, Haberlandt was proposing the administration of hormones for birth control. An extract was produced, named infecundin, ready to be used, but Haberlandt's early death in 1932, at age 47, brought an end to this effort. Fellner disappeared after the annexation of Austria to Hitler's Germany.

The concept was announced 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 an eccentric chemist, Russell E. Marker, who completed his thesis, but not his course work, for his Ph.D. Marker, born in 1902 near Hagerstown, Maryland, received his Bachelor's degree in organic chemistry and his Master's degree in colloidal chemistry from the University of Maryland. After leaving the University of Maryland, Marker worked with the Ethyl Gasoline Corporation, and in 1926, developed the process of octane rating, based on the discovery that knocking in gasoline was due to hydrocarbons with an uneven number of carbons.

From 1927 to 1935, 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 solving the problem of producing abundant and cheap amounts of progesterone, but he was told to continue with his work in optical technology. In 1935, he moved to Pennsylvania State University at a reduced salary, but with the freedom to pursue any field of research. At that time, it required the ovaries from 2500 pregnant pigs to produce 1 mg of progesterone. In 1939, Marker devised the method (called the Marker degradation) to convert a sapogenin molecule into a progestin. Marker became 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.

This conviction was strengthened with his discovery that a species of Trillium, known locally as Beth's root, was collected in North Carolina and used in the preparation of Lydia Pinkham's Compound, popular at the time to relieve menstrual troubles. The active ingredient in Beth's root was diosgenin, but the thimzone was too small to provide sufficient amounts for commercial production. Marker's search for an appropriate plant took him to California, Arizona, and Texas.

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 waiting for his wife to return home. On the night of a retired botanist. After returning to Pennsylvania, he decided to go to Veracruz, Mexico (it took 3 days by train), to search this dioscorea. He made several attempts in 1941 and early 1942, but was frustrated first by the lack of a plant-collecting permit from the Mexican government and then by his failure to find the plant. He remembered that the book with the picture reported that this dioscorea was known locally as “cabeza de negro,” black tubers that grew near Orizaba and Cordoba. Marker took a bus to Cordoba, and near Orizaba, an Indian who owned a small store brought him two plants. Each tuber was 9–12 inches high and consisted of white material like a turnip, used by local Mexicans as a poison to catch fish.

Marker managed to get one tuber of brown tubers back to Pennsylvania State University and isolated diosgenin. Unable to obtain support from the pharmaceutical industry, Marker used his life savings, and in 1942, he returned to Veracruz, collected the roots of the Mexican yam, and prepared a syrup from the roots. Back in Pennsylvania with his 5-gallon cans of syrup, Marker worked out the degradation of diosgenin to progesterone. One 5-gallon can yielded 3 kg of progesterone. United States pharmaceutical companies still refused to back Marker, and even the University refused, despite Marker's urging, to patent the process.

In 1943, Marker resigned from Pennsylvania State University and went to Mexico where he collected the roots of Dioscorea mexicana, 10 tons worth! Looking through the yellow pages in a Mexico City telephone directory, Marker found a company called Laboratorios Hormona, owned by a lawyer, Emeric Somlo, and a physician, Frederick Lehman. Marker arranged a meeting, and the three agreed to form a Mexican company to produce hormones. In an old pottery shed in Mexico City (the laboratories of Laboratorios Hormona), in two months, he prepared several pounds of progesterone (worth $300,000) with the help of four young women who had little education and spoke no English (Marker did not speak Spanish). The two partners and Marker formed a company in 1944 that they called Syntex (from synthesis and Mexico). In 1944, Marker produced over 30 kg of progesterone. The price of progesterone fell from $200 to $50 a gram.

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. Failing to reach a settlement, Marker left Syntex after only one year 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 $1 a gram, and later to $5. This company was allegedly harassed (legally and physically) by Syntex, and in 1946, sold, eventually reaching ownership by Organon of Holland, which still uses it.

In 1949, Marker retired to Pennsylvania to devote the rest of his life to making replicas of antique works in silver, a successful business that allowed him, in the 1980s, to endow scientific lectureships at both Pennsylvania State University and the University of Maryland. However, he took his know-how with him. Fortunately for Syntex, he had published a scientific description of his process, and there still was no patent on his discoveries. Syntex recruited George Rosenkranz, a Hungarian immigrant living in Cuba, to reinstitute the commercial manufacture of progesterone (and testosterone) from Mexican yams, a task that took him (with the help of the women left behind by Marker) 2 years.

In 1970, he formed the Mexican government recognized Marker and awarded him the Order of the Aztec Eagle; 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. At the age of 92, Russell Earl Marker died in Wernersville, Pennsylvania, in 1995, from complications after a broken hip.

**Carl Djerassi**

The Djerassi family lived in Bulgaria for hundreds of years after escaping Spain during the Inquisition. Carl Djerassi, the son of a Bulgarian physician, was born in Vienna (as was his physician mother). Djerassi, at the age of 16, and his mother emigrated to the United States in 1939. A Jewish refugee aid organization placed Djerassi with a family in New York, 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.

In 1949, it was discovered that cortisol relieved arthritis, and the race was on to develop an easy and cheap method to synthesize cortisol. Djerassi, at age 26, joined Syntex to work on this synthesis. At the Mexican yam plant, while diosgenin was the starting point. This was quickly achieved (in 1951), but soon after, an even better method of cortisone production using microbiologic fermentation was discovered at Upjohn. This latter method used progesterone as the starting point, and, therefore, Syntex found itself as the key supplier to other companies for this important process, at the rate of 10 tons of progesterone per year and a price of 48 cents per gram.

Djerassi and other Syntex chemists then turned their attention to the sex steroids. They discovered that the removal of the 19-carbon from yam-derived progesterone increased the prostaglandinal activity of the molecule. Ethisterone had been available for a dozen years, and the Syntex chemists reasoned that removal of the 19-carbon would increase the prostaglandinal potency of this orally active compound. In 1951, norethindrone was synthesized; the patent for this drug is the first patent for a drug listed in the National Inventor's Hall of Fame in Akron, Ohio. A closely related compound, norethynodrel, was actually the first orally active prostaglandinal agent to receive a patent, assigned to Frank Colton, a chemist at G.D. Searle & Company.

**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 German-Jewish philanthropic organization. Pincus was the oldest of 6 children and grew up in a home of intellectual curiosity and energy, but even his family regarded him as a
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 genius. By 1936, Harvard had cited Pincus's work as one of the university's outstanding scientific achievements of all time, but Harvard denied him tenure 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 tenure 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 received his Ph.D. degree from Harvard on an infamous day, December 7, 1941, and thus he was forced to remain in this country. He was drawn to Pincus because of Pincus's book, The Eggs of Mammals, published in 1936, a book that had a major impact on biologists at that time. The successful recruitment of M-C Chang by Hoagland and Pincus was to pay great dividends.

Soon Hoagland had put together a group of outstanding scientists, but because of his on-going 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, 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. The Board of Trustees was chaired by Harlow Shapley, a distinguished astronomer, vice-chaired by Rabbi Levi Olan, and included 3 Nobel laureates and a group of Worcester businessmen. 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 3000 scientific papers were published.

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.

The initial discoveries that led to an oral contraceptive can be attributed to M-C Chang (also the first to describe the capacitation process of sperm). In 1951, he confirmed the work of Makepeace (in 1937) demonstrating that progesterone could inhibit ovulation in rabbits. When norethindrone and norethynodrel became available, Chang found them to be virtually 100% effective in inhibiting ovulation when administered orally to rabbits.

Katherine Dexter McCormick was a very rich woman; in 1904, she married Stanley McCormick, the son of Cyrus McCormick, the founder of International Harvester. She was also intelligent, the second woman to graduate from the Massachusetts Institute of Technology, socially conscious, and a generous contributor to family planning efforts. McCormick'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.

Pincus attributed his interest in contraception to his growing appreciation for the world's population problem, and to a 1951 visit with Margaret Sanger, at that time president of the Planned Parenthood Federation of America. At that visit, Sanger expressed hope that a method of contraception could be derived from the laboratory work being done by Pincus and Chang.

In 1952, Margaret Sanger brought Pincus and Katherine McCormick together. During this meeting, Pincus formulated his thoughts derived from his mammalian research. He envisioned a postgestational agent in pill form as a contraceptive, acting like progesterone in pregnancy. Sanger and McCormick provided a research grant for further animal research. By the time of her death, McCormick had contributed more than $2 million to the Worcester Foundation, and left another $1 million in her will. In his book, The Control of Fertility, published in 1965, Pincus wrote: "This book is dedicated to Mrs. Stanley McCormick because of her steadfast faith in scientific inquiry and her unswerving encouragement of human dignity."

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, probably the first demonstration of fertilization of human oocytes in vitro. Rock was interested in the work with postgestational agents, not for contraception however, but because he hoped the female sex steroids could be used to overcome 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 (pleasing Rock who all along was motivated to treat infertility).

Pincus and Chang decided to announce their findings at the International Planned Parenthood meeting in Tokyo, in the fall of 1955. Rock refused to join in this effort, believing that Pincus and Chang were moving too fast. Despite this disagreement (which apparently was spirited and strong), it was done, and the Tokyo presentation generated worldwide publicity.

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 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 accelerated licensing of contraception. In 1957, Enovid was approved for the treatment of miscarriages and menstrual disorders, and in 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."

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, 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 Katherine McCormick at the age of 92), of aplastic anemia that some have argued was caused by his long-term exposure to solvents and chemicals. Rock died in 1964, at the age of 94, and Chang, in 1991, was buried at the age of 82, in Shrewsbury, near his laboratory and close to the grave of Pincus. Pincus wrote his book, The Control of Fertility, in 1964–65, only because “a break came in the apparent dam to publication on reproductive physiology and particularly its subdivisions concerned with reproductive behavior, conception, and contraception.”

“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.”

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. As a result, 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 inactivity 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 made estradiol orally active. Ethinyl estradiol is a very potent oral estrogen and is one of the two forms of estrogen in every oral contraceptive. The other estrogen is 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 have 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 all of the low-dose oral contraceptives contain ethinyl estradiol.

The metabolism of ethinyl estradiol (particularly as reflected in blood levels) varies significantly from individual to individual, and from one population to another. 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 none in another.

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.

The “impurity” of 19-nortestosterone, i.e., androgenic as well as progesteronal 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. 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. Clinically, androgenic and estrogenic activities of the progestin component are, therefore, 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, not the particular progestin, 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.

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 undergoes two metabolic steps before the progestational activity is expressed in its active metabolite, 3-keto-desogestrel. 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 D-15 gestodene. It is metabolized into many derivatives with progestational activity, but not levonorgestrel. Several metabolites contribute to the activity of norgestimate, including 17-deacetylated norgestimate, 3-keto norgestimate, and levonorgestrel. Although norgestimate is a "new" progestin, epidemiologists have included it in the oral contraceptive second generation family because its activity is believed to be largely due to levonorgestrel and levonorgestrel metabolites, although this may not be totally accurate. 5 10

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 30 or 35 µg ethinyl estradiol

Third Generation Oral Contraceptives — Products containing desogestrel or gestodene with 20 or 30 µg ethinyl estradiol

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.

**Potency**

For many years, clinicians, scientists, medical writers, and even the pharmaceutical industry have 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 upon 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. Potency is no longer an important clinical issue.

**New Progestins**

Probably the greatest influence on the effort that yielded the new progestins was the belief throughout the 1980s that anabolic 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. In the search to find compounds that minimize anabolic effects, however, the pharmaceutical companies succeeded.

The new progestins include desogestrel, gestodene, and norgestimate, and even newer progestins are in development. 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 new 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 may be of greater clinical value in the treatment of acne and hirsutism, but appropriate comparative clinical studies to document a better response have not been performed.

The new progestins, because of their reduced androgenicity, predictably do not adversely affect the cholesterol-lipoprotein profile. Indeed, the estrogen-progesterin balance of combined oral contraceptives containing one of the new progestins may 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 will be difficult to accumulate data with these rare events.

**New 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. We are probably at or very near the lowest dose levels that can be achieved without sacrificing efficacy. Metabolic studies with the multiphasic preparations indicate no differences or slight improvements over the metabolic effects of low-dose monophasic products.

**Mechanism of Action**

The combination pill, consisting of estrogen and progestin components, is given daily for 3 of every 4 weeks. The combination pill 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 selection and 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-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 cervical mucus, and perhaps tubal function reflect progestational stimulation. Even if there is some ovarian follicular activity (especially with the lowest dose products), these actions serve to ensure good contraceptive efficacy.

**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 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. The most prevalent problems that can be identified associated with apparent oral contraceptive failures are vomiting and diarrhea. 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.

The contraceptive effectiveness of the new low-dose 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. While carefully monitored studies with motivated subjects achieve an annual failure rate of 0.1%, typical usage is associated with a 3.0% failure rate during the first year of use. Efficacy decreases slightly when the estrogen component is removed, and only a small dose of the progestin is administered (the progestin-only minipills).
Levonorgestrel IUD 0.6 0.8
Copper T 380A 0.1 0.1
Norplant 0.05 0.05
Female sterilization 0.05 0.05
Male sterilization 0.1 0.15
Depo-Provera 0.3 0.3
Spermicides 6.0 26.0
Periodic abstinence
Calendar 9.0 25.0
Ovulation method 3.0
Symptothermal 2.0
Post-ovulation 1.0
Withdrawal 4.0 19.0
Cervical cap
Parous women 26.0 40.0
Nulliparous women 9.0 20.0
Sponge
Parous women 9.0 28.0
Nulliparous women 6.0 18.0
Diaphragm and spermicides 6.0 20.0
Condom
Male 3.0 14.0
Female 5.0 21.0

Metabolic Effects of Oral Contraception

Cardiovascular Disease

In October, 1995, the United Kingdom Committee on Safety of Medicines sent a letter to all U.K. physicians and pharmacists stating that women taking oral contraceptives containing desogestrel or gestodene should be urged to complete their current cycle and to continue a formulation with these progestins only if prepared to accept an increased risk of venous thromboembolism. The Committee on Safety of Medicines took this action because of observational studies that indicated a two-fold increase in the risk of venous thromboembolism when desogestrel- and gestodene-containing contraceptives were compared with products with other progestins (mostly levonorgestrel). This action and the studies on which it was based immediately became controversial. The controversy went beyond the validity of the epidemiologic data. The publicity surrounding these events reverberated throughout Europe, leading to an immediate overall decrease in oral contraceptive use, an increase in unwanted pregnancies, and an increase in induced abortions. 17, 18

The controversy involving new progestin oral contraceptives that began in late 1995, continued through 1996, and began to reach resolution in 1997. The fundamental question is whether oral contraceptives containing desogestrel and gestodene have a different risk of thrombosis compared with oral contraceptives containing older progestins. 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 Coagulation System

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 Xla. 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. A mutation in the prothrombin gene and the factor V Leiden mutation are the most common inherited causes of venous thromboembolism. 19

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. 20 The factor V Leiden mutation is found in approximately 30% of individuals who develop venous thromboembolism. 20 Activated protein C inhibits coagulation by degrading factors V and VIII. One of the 3 cleavages 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). 20 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.

Coagulation and Fibrinolysis Factors

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
Plasminogen activator inhibitor-1 (PAI-1)

Factors that favor clotting when decreased
Antiplasmin

Heterozygotes for the factor V Leiden mutation have an 8-fold increased risk of venous thrombosis, and homozygotes, an 80-fold increased risk, and this risk appears
to be 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. The mutation is believed to have arisen in a single ancestor approximately 21,000 to 34,000 years ago. 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. The prevalence of this abnormality in the white population is estimated to range from 0.7% to 4%.

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. 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. 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. However, these changes are essentially all within normal ranges and their clinical significance is unknown. Smoking produces a shift to hypercoagulability. A 20 µg estrogen formulation has been reported to have no effect on clotting parameters, even in smokers. One study comparing a 20 µg product with a 30 µg product found similar mild pro-coagulant and fibrinolytic activity, although there was a trend toward increased fibrinolytic activity with the lower dose.

There is no evidence of an increase in risk of cardiovascular disease among past users of oral contraception. In the Nurses' Health Study and the Royal College of General Practitioners' Study, long-term past use of oral contraceptives was not associated with an increase in overall mortality. 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 — The Conventional Wisdom

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. 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. Smoking was documented to produce an additive increase in the risk of arterial thrombosis, but had no effect on the risk of venous thromboembolism.

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. 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 Pune Sound study in the United States documented a reduction in venous thrombosis risk to 2-fold. The new studies also reflect the importance of these two forces, but they still indicate an increased risk.

The Hierarchy of Epidemiologic Studies

I. Clinical Reports

A case report: An anecdotal report that serves to bring attention to a possible problem or condition.

A case series: A collection of similar cases that suggests more than a chance or coincidental occurrence.

II. Observational Studies (Non-experimental Studies: Observation Without Intervention)

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

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

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.

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 surveillance bias).

Example: The Nurses' Health Study

III. Randomized Controlled Trials

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.

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
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 oral 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 and those who are not, 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 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.

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 risks 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. 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. 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.

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.

Venous Thromboembolism — The Controversial Studies

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. 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 µg 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). There were only 9 cases and 3 controls using combined oral contraceptives with other progestins, precluding precise analysis. The users of the levonorgestrel formulation had an increased odds ratio (an estimation of relative risk used in case-control studies) 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. Also of note, the increased risk for the desogestrel formulation containing 20 µg ethinyl estradiol was 38.2, a number that is obviously not reliable because it was based upon only 8 cases and 1 control; the confidence interval (CI) of 4.5–325 reflected this imprecision. Overall, these increased risks were lower than those estimated by earlier case-control studies of higher dose oral contraceptives.

The second case-control study (from an international team of epidemiologists and called 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. 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 (CI = 2.3–4.3). Comparing users of desogestrel and gestodene products to users of second generation oral contraceptives, the risk of venous thromboembolism was 1.5-fold greater.

The third 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. Using this cohort, the authors calculated the death rate from pulmonary embolism, stroke, and acute myocardial infarction 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. There were only 4 cases and 9 controls using the 20 µg ethinyl estradiol and desogestrel product, and although the risk was similar to that associated with the 30 µg ethinyl estradiol and desogestrel product, this is too small a number for analysis.

Similar results were reported when women with deep vein thrombosis in the Leiden Thrombophilia Study in the Netherlands were re-analyzed for their use of oral contraceptives. 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.

Smoking, well recognized as a risk factor for arterial thrombosis, did not affect the risk estimates in these studies. This is not a new observation; older studies of venous thromboembolism also failed to identify smoking as a risk factor.

Venous Thromboembolism — Subsequent Studies

The publication of the 4 reports in late 1995 and early 1996 was followed by a flood of letters to editors, as well as reviews and editorials, highlighting confounding and bias problems in these studies. Some prominent figures were convinced the reports of increased risks with desogestrel and gestodene were real;
others were skeptical, pointing out possible confounding biases. Subsequently, re-analysis and new studies revealed confounders and biases in the initial studies. Thus, a consistent picture gradually emerged with consideration of proper analysis of the generated data, and the adjustment for confounding biases not initially apparent.

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). 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 a significant difference 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.

A case-control study using 83 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 below] 51). In this study, matching cases and controls by exact 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. Thus, in these two studies, more precise adjustments for age eliminated a confounding bias.

A re-analysis of the Transnational Case-Control Study considered the duration and patterns of oral contraceptive use. This re-analysis 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.

Evaluation of the Studies

An immediate problem with the initial studies was how to reconcile the results with the conventional wisdom that thrombosis is an estrogen dose-related complication. Furthermore, prostaglandins, and desogestrel and gestodene in particular, have no significant impact on clotting parameters. Therefore, there was inherent biologic implausibility surrounding the new studies.

The initial reports resurrected the claim by Kuhl in 1988 and 1989 that gestodene could cause more thrombosis because it affected ethinyl estradiol metabolism, resulting in higher estrogen levels. Other laboratories, however, could not replicate Kuhl's findings.

Former users discontinue oral contraceptives for a variety of reasons, and often are switched to what clinicians perceive to be "safer" products ("preferential prescribing") 51, 52 and 53. 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 ("healthy user effect"). This is also called attrition of susceptibles because higher risk individuals with problems are gradually eliminated from the group. 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. 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 a congenital predisposition to venous thrombosis). These changing practice patterns exert different effects over the lifetime of a product, and analytical adjustments are extremely difficult. The Transnational Group believed it accomplished an appropriate adjustment by focusing on first-time users and duration of use. It is also unlikely that the "healthy user effect" will be dominant in first-time users. And, of course, this analysis found no differences between second and third generation oral contraceptives.

The challenge for a clinician is to make a decision: is an observational study with statistically significant results clinically (biologically) real? This controversy illustrates how difficult this can be. When faced with results from observational studies, clinicians want to see uniformity, consistency, agreement—all arguing in favor of a real clinical effect. Examples are the protective effect of oral contraceptives on the risk of ovarian cancer, and the benefits of postmenopausal estrogen therapy on cardiovascular disease. 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.

The apparent differences associated with the new progestins, it is now apparent, were due to the marketing and preferential prescribing of new products, which influenced the characteristics of the patients for whom the new products were prescribed. Most impressive and important is the fact that there is no evidence of an increase in mortality due to venous thromboembolism since the introduction of new progestin oral contraceptives.

Venous Thromboembolism and the Factor V Leiden Mutation

The new studies indicate that a risk of idiopathic venous thrombosis persists with low-dose oral contraceptives, at a level of approximately 3–4-fold greater than the normal, general incidence. Heterozygotes have a 6–8-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. There is no known association between the factor V Leiden mutation and arterial thrombosis.

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: Caucasian Americans 5.27%
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 in women who use oral contraceptives is increased only in smokers. Furthermore, because only a small number of women even with the Leiden mutation (less than 1 in 1000) 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 (which has an even greater risk of venous thromboembolism) would likely follow. 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.

This aspect of the oral contraceptive venous thromboembolism controversy received a transfusion of energy with the publication of a report from the Netherlands, utilizing a laboratory test for resistance to activated protein C to compare differences in oral contraceptive non-users, users of second generation oral contraceptives, and users of third generation oral contraceptives. Women who used any oral contraceptive had a decreased sensitivity to activated protein C compared with nonusers, and women who used third generation oral contraceptives were even less sensitive, and less than users of second generation products. This was presented as an explanation for the epidemiologic data indicating greater risks of venous thromboembolism associated with desogestrel and gestodene, and an editorial in the April 19, 1997, issue of Lancet concluded that this report was the “nail in the coffin” confirming the epidemiologic evidence.

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. 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 United Kingdom, the results were as follows:

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any OC use</td>
<td>57</td>
<td>156</td>
<td>2.35</td>
</tr>
<tr>
<td>50 µg estrogen OCs</td>
<td>14</td>
<td>22</td>
<td>4.32</td>
</tr>
<tr>
<td>Old progestin OCs</td>
<td>28</td>
<td>71</td>
<td>2.96</td>
</tr>
<tr>
<td>New progestin OCs</td>
<td>7</td>
<td>49</td>
<td>0.82</td>
</tr>
</tbody>
</table>

These data were interpreted as indicating no increased risk of myocardial infarction associated with oral contraceptives containing desogestrel or gestodene. However, the reduced risk with the new progestin oral contraceptives was also emphasized (the comparison of third generation products to second generation products yielded a reduced risk that was statistically significant), suggesting a possible saving of deaths from myocardial infarction with desogestrel and gestodene. The problem is that the small actual incidence makes it difficult to acquire sufficient numbers. The conclusion was based on only 7 cases and 49 controls using third generation oral contraceptives and 28 cases and 71 controls using second generation products, and, in our view, the power is too limited to make any conclusion regarding the new progestin oral contraceptives. This is a good example of a conclusion that may be statistically significant, but clinically not real.

Atherosclerosis — Myocardial Infarction

Many different factors contribute to cardiovascular disease (especially hypertension, migraine with aura, diabetes mellitus, and smoking).

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, stroke is the most important possible side effect. A very low incidence of stroke in young women carries with it little increase in absolute risk. However, because the incidence of cerebral thrombotic attacks is higher in women over age 40, we should do our best, as the following paragraphs will indicate, to make sure oral contraceptive users over age 40 are in good health and without significant risk factors for cardiovascular disease (especially hypertension, migraine with aura, diabetes mellitus, and smoking).

It has been difficult to establish arterial thrombosis dose-response relationships with estrogen because these events are so rare. Nevertheless, the estrogen dose is important for the risk of myocardial infarction and thrombotic strokes. Thus, a rationale for advocating low-dose estrogen oral contraceptives continues to be valid. It is acknowledged that the apparent higher risk of venous thromboembolism associated with 20 µg products in the WHO and Transnational studies reflects small numbers as well as preferential prescribing and healthy user effects.

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. 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 United Kingdom, the results were as follows:

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any OC use</td>
<td>57</td>
<td>156</td>
<td>2.35</td>
</tr>
<tr>
<td>50 µg estrogen OCs</td>
<td>14</td>
<td>22</td>
<td>4.32</td>
</tr>
<tr>
<td>Old progestin OCs</td>
<td>28</td>
<td>71</td>
<td>2.96</td>
</tr>
<tr>
<td>New progestin OCs</td>
<td>7</td>
<td>49</td>
<td>0.82</td>
</tr>
</tbody>
</table>

These data were interpreted as indicating no increased risk of myocardial infarction associated with oral contraceptives containing desogestrel or gestodene. However, the reduced risk with the new progestin oral contraceptives was also emphasized (the comparison of third generation products to second generation products yielded a reduced risk that was statistically significant), suggesting a possible saving of deaths from myocardial infarction with desogestrel and gestodene. The problem is that the small actual incidence makes it difficult to acquire sufficient numbers. The conclusion was based on only 7 cases and 49 controls using third generation oral contraceptives and 28 cases and 71 controls using second generation products, and, in our view, the power is too limited to make any conclusion regarding the new progestin oral contraceptives. This is a good example of a conclusion that may be statistically significant, but clinically not real.

The Transnational study found that cigarette smoking carried a higher risk for myocardial infarction than oral contraceptives, and that nonsmoking users of oral contraceptives had no evidence of an increased risk. In addition, there was an indication that patient screening is important in minimizing the impact of hypertension on the risk of myocardial infarction.

In the WHO multicenter study, there were 368 cases of acute myocardial infarction. 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 is than 35 years old. There was no apparent relationship with estrogen dose, and there was no apparent influence of type or dose of progestin. However, 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 Dutch 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. 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.
Arterial Thrombosis — Stroke

Older case-control and cohort studies indicated an increased risk of cerebral thrombosis among current users of high-dose oral contraceptives. 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. 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 two-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.  

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. 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 also analyzed their data for ischemic stroke in a case-control study of 220 ischemic strokes in the United Kingdom, Germany, France, Switzerland, and Austria. 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 prostogens. 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.

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.

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. 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 1968 cases. 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. In this study, the 219 cases during 1994 and 1995 included 146 cases of cerebral infarction and 73 cases of transient ischemic attacks. Only users of 2 generation oral contraceptives (levonorgestrel, norgestrel, and norgestimate) had a statistically significant increased risk (about 2.5-fold). 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. Risk was increased with smoking, treated hypertension, diabetes, heart diseases, frequent migraine, a family history of myocardial infarction, but not duration of use, or family history of venous thromboembolism.

### Incidence of Stroke in Reproductive Age Women

<table>
<thead>
<tr>
<th>Incidence of ischemic stroke</th>
<th>5 per 100,000 per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–3 per 100,000 per year</td>
<td>in women under age 35</td>
</tr>
<tr>
<td>10 per 100,000 per year</td>
<td>in women over age 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incidence of hemorrhagic stroke</th>
<th>6 per 100,000 per year</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Excess cases per year due to OCs, including smokers and hypertensives</th>
<th>2 per 100,000 per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 per 100,000 per year in low-dose OC users</td>
<td></td>
</tr>
<tr>
<td>8 per 100,000 per year in high-dose OC users</td>
<td></td>
</tr>
</tbody>
</table>

Arterial Thrombosis — Current Assessment

There has been no evidence with respectable statistical power that the new progestins have an appreciable difference in risk for arterial disease, an event that is
already NOT increased with low-dose older type progestin oral contraceptives. It is possible that as these studies continue and acquire greater statistical power, a difference will emerge, but even if this is the case, the difference will be minor and likely unmeasurable. Conclusions based on a limited number of cases are premature, and a critical attitude toward arterial thrombosis is appropriate just as such an approach finally revealed explanations for the initial findings with venous thrombosis.

Most importantly, the new 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.

Over the years, there has been recurring discussion over whether to provide oral contraceptives over the counter on a non-prescription basis. The data in the WHO report 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 non-clinical source. Deprived of screening, women with risk factors in developing countries were exposed to greater risk.

**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. The new studies indicate that hypertension should be a major concern, especially in regards to the risk of stroke. Certainly, women with uncontrolled hypertension should not use oral contraceptives. Generally, family planning experts have believed that well-treated hypertension should not be a contraindication for oral contraceptive use. The new data do not help us with this problem because it is impossible to accurately categorize hypertensive patients in the studies into groups representing successful and unsuccessful treatment. Nevertheless, the outstanding safety of low estrogen dose oral contraceptives in these studies supports the continued use of low-dose oral contraceptives in treated and well-controlled hypertensive women.

**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 compared to 38.5% of a similar group of nonusers. Prolonged oral contraceptives, with or without smoking, appear not to alter smoking habit. Despite the demonstrated morbidity and mortality from smoking, both health education and publicity that smoking is a contraindication to oral contraceptive use. The new data do not help us with this problem because it is impossible to accurately categorize smokers in the studies into groups representing successful and unsuccessful treatment. Nevertheless, the outstanding safety of low estrogen dose oral contraceptives in these studies supports the continued use of low-dose oral contraceptives in treated and well-controlled hypertensive women.

**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. The levonorgestrel triphasic exerts no significant changes on HDL-cholesterol, LDL-cholesterol, apoprotein B, and no change 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. Like the triphasic levonorgestrel pills, norethindrone multiphasic pills have no significant impact on the lipoprotein profile over 6–12 months. 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.

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 arteriosclerosis despite a reduction in HDL-cholesterol levels. 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 arteriosclerosis than nonsmokers. 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.

In the past decade, we have been 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).

**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 mg estrogen, monophasic pills, including those containing the new progestins. However, an increased incidence of clinically significant hypertension has not been reported. 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. 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. 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. Likewise, women who have developed hypertension on oral contraception are not more predisposed to develop toxemia of pregnancy.

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. In our view, with medical control of the blood pressure and close follow-up (at least every 3 months), the patient and her clinician may choose low-dose oral contraception. Close follow-up is also indicated in women with a history of congestive renal disease or a strong family history of hypertension or cardiovascular disease. It seems prudent to suggest that patients with
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 several studies. Although the risk of venous thromboembolism is slightly increased, the actual incidence is still relatively rare, and the mortality rate 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.

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. 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.

<table>
<thead>
<tr>
<th>Hypercoaguable Conditions</th>
<th>Thrombophilia Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antithrombin III deficiency</td>
<td>Protein C III</td>
</tr>
<tr>
<td>Protein C deficiency</td>
<td>Protein C</td>
</tr>
<tr>
<td>Protein S deficiency</td>
<td>Protein S</td>
</tr>
<tr>
<td>Factor V Leiden mutation</td>
<td>Activated protein C resistance ratio</td>
</tr>
<tr>
<td>Prothrombin gene mutation</td>
<td>Activated partial thromboplastin time</td>
</tr>
<tr>
<td>Antiphospholipid syndrome</td>
<td>Hexagonal activated partial thromboplastin time</td>
</tr>
<tr>
<td></td>
<td>Anticardiolipin antibodies</td>
</tr>
<tr>
<td></td>
<td>Lupus anticoagulant</td>
</tr>
<tr>
<td></td>
<td>Fibrinogen</td>
</tr>
<tr>
<td></td>
<td>Prothrombin G mutation (DNA test)</td>
</tr>
<tr>
<td></td>
<td>Thrombin Time</td>
</tr>
<tr>
<td></td>
<td>Homocysteine level</td>
</tr>
<tr>
<td></td>
<td>Complete blood count</td>
</tr>
</tbody>
</table>

Combination oral 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. 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 seems 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.

The conclusion once again is that low-dose oral contraceptives are very safe for healthy, young women. By effectively screening for the presence of smoking and cardiovascular risk factors, especially hypertension, in older women, 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 (long-term) use.

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-hour values may be slightly elevated.

Insulin sensitivity is affected mainly by the progestin component of the pill. 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. 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. The reason this is important is that it is now...
recognized that hyperinsulinemia due to insulin resistance is a contributor to cardiovascular disease. However, there are several critical questions that remain unanswered. Can the results from a cross-sectional study be duplicated in a study of sufficient size with patients serving as their own controls? Is a statistically significant hyperinsulinemia, detected in a study, clinically meaningful?

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). 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. The 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. 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. According to the older epidemiologic data, the use of oral contraceptives increases 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 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 found no evidence that oral contraceptive use by young women with insulin-dependent diabetes mellitus increased the development of retinopathy or nephropathy. 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 hematostatic biochemical markers for cardiovascular risk.

The Liver
The liver is affected in more ways and with more regularity and intensity by the sex steroids than by 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.

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 oral contraceptive use is acute or chronic hepatic failure. Cirrhosis and previous hepatitis are not aggravated. Once recovered from the acute phase of liver disease, a woman can use oral 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. 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 the determining factor. The Nurses’ Health Study reported no significant increase in risk of symptomatic gallstones among ever-users, but slightly elevated risks among current and long-term users. 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. 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.

Other Metabolic Effects
Nausea, breast discomfort, and weight gain 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. 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 fail to indicate a difference in body weight between users and nonusers.

There is no association between oral contraception and peptic ulcer disease or inflammatory bowel disease.

Chloasma, a patchy increase in facial pigmentation, was, at one time, found to occur in approximately 5% 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.

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 continuous 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 (B6) 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.

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.

Adverse androgenic voice changes were occasionally encountered with the use of the first very high-dose oral contraceptives. Vocal vibrillation 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 now a side effect of concern.

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. 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. This protection is equally protective for all 3 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. There are no data as yet with multiphasic preparations or the new progestin...
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. 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. This protective effect increases with duration of use and continues for at least 10–15 years after stopping the medication. 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), reaches 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. 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). 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. 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 a case-control study of women with BRCA1 or BRCA2 mutations.

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 one year. Invasive cervical cancer may be increased for 5 years of use, reaching a two-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 the most 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 was 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). 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.

A case-control study of patients in Panama, Costa Rica, Colombia, and Mexico concluded that there was a significantly increased risk for invasive adenocarcinoma. Similar results were obtained in a case-control study in Los Angeles and in the World Health Organization Collaborative Study. 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. 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. Oral contraceptives increase cervical ectopia, but whether this increases the risk of cervical adenocarcinoma is unclear.

This concern obviously is an important reason for annual Pap smear surveillance. 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 diseases). Oral contraceptive use is appropriate for women with a history of cervical intraepithelial neoplasia (CIN), including those who have been surgically treated.

Liver Adenomas

Hepatocellular adenomas can be produced by steroids of both the estrogen and androgen families. Actually, there are several different lesions, poliosis, focal nodular hyperplasia, and adenomas. Poliosis is characterized by dilated vascular spaces without endothelial lining, and may occur in the absence of adenomatous changes. The adenomas are not malignant; their signifying lies in the potential for hemorrhage. The most common presentation is acute right upper quadrant or epigastric pain. The tumors may be asymptomatic, or may present suddenly with hematoperitoneum. There is some evidence that the tumors and focal nodular hyperplasia regress when oral contraception is stopped. 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. In our view it is not even worth mentioning during the informed consent (choice) process.

No reliable screening test or procedure is currently available. 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.

Liver Cancer

Oral contraception has been linked to the development of hepatocellular carcinoma. However, the very small number of cases, and, thus, the limited statistical power, requires great care 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. 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. In the United States, Japan, Sweden, England, and Wales, the death rates from liver cancer have not changed over the last 3 decades despite introduction and use of oral contraception.

Breast Cancer

Because of its prevalence and its long latent period, concern over the relationship between oral contraception and breast cancer continues to be an issue in the minds of both patients and clinicians. Unfortunately, the issue is not totally resolved and probably will not be until another decade passes, allowing data to emerge from the modern era of lower dose oral contraception.

Over the last decade, 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 cohort of women who have used oral contraception in this fashion is just now beginning to reach the ages of postmenopausal breast cancer, the 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, while others indicated no increase in overall risk. The most impressive finding indicates a link in most studies, but not all, of early breast cancer before age 40 with women who used oral contraception for long durations of time.

A collaborative group composed of an enormous number of epidemiologists and cancer investigators from around the world 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. Oral contraceptives were grouped into 3 categories: low, medium, and high dose (which correlated with <50µg, 50 µg, and >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 was based not at all and 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 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: Relative Risk = 0.88, CI = 0.81–0.95.

Oral Contraceptives and the Risk of Breast Cancer
Re-analysis of the World's Data

<table>
<thead>
<tr>
<th>Current users</th>
<th>RR = 1.24, 95% CI 1.15–1.33</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–4 years after stopping</td>
<td>RR = 1.16, 95% CI 1.08–1.23</td>
</tr>
<tr>
<td>5–9 years after stopping</td>
<td>RR = 1.07, 95% CI 1.02–1.13</td>
</tr>
</tbody>
</table>

Data were limited for progestin-only methods. The re-analysis 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 yield good statistical significance.

Overall, this massive statistical exercise yielded good results. 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.

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. Recent 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. And some have found that a concurrent event or recent pregnancy transiently increases the risk of breast cancer (as reviewed in Chapter 15). 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 contraceptives. In women under 50 years of age, there was only a slight increase during this same time period.

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 15) on premenopausal breast cancer, the cancer of concern to younger women using oral contraception.

Conclusion

Adding up the benefits of oral contraception, the possible slight increase in risk of breast cancer is far outweighed by positive effects on our public health. But the impact on public health is of little concern during the private clinician—patient interchange in the office. Here personal risk receives highest priority; fear of cancer is a motivating force, and compliance with effective contraception requires accurate information. For these reasons, we provide the following summary of our assessment of the impact of oral contraceptives on the risk of breast cancer.

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 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 15). 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 contraceptives. In women under 50 years of age, there was only a slight increase during this same time period.
- 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 15) on premenopausal breast cancer, the cancer of concern to younger women using oral contraception.

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 recent and 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. There is no evidence linking oral contraceptive use to kidney cancer, gallbladder cancer, or pituitary tumors. Long-term oral contraceptive use may slightly increase the risk of molar pregnancy, but there is no convincing evidence of a cause-and-effect association. Although previous studies have not been in agreement, the Nurses' Health Study reports about a 40% reduced risk of colorectal cancer associated with 8 years of previous use of oral contraceptives (most likely higher dose products).

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. 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 11b-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.
Thyroid

Estrogen increases the synthesis and circulating levels of thyroxine-binding globulin. Prior to the introduction of new 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.

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 deserve 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 LMP).

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. However, subsequent analysis of these data uncovered several methodologic shortcomings. Simpson, in a very thorough and critical review in 1990, concluded that there was no reliable evidence implicating sex steroids as cardiac teratogens. 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 which 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. A meta-analysis of 26 prospective 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.

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 women who have been exposed to a gestational agent such as medroxyprogesterone acetate or 17-hydroxyprogesterone caproate.

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. Presumably, the delay is 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 have never given birth. 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. It should be noted that after 72 months the proportions of women who remained unchilded were the same in both groups of women.

This delay has been observed in the United States 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 users, and 11.9% for other methods). 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 indicate that this delay lasts at least a year, and the effect is greater with higher dose preparations. Despite this 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.

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. 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.

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. 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). 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 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. Former pill users have no increased risks for the following: perinatal morbidity or mortality, prematurity, and low birth weight. Dizygous twinning has been observed to be nearly two-fold (1.6% versus 1.0%) increased in women who conceive soon after cessation of oral contraception. 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.
Breastfeeding

Oral contraception has been demonstrated to diminish the quantity and quality of lactation in postpartum women. Also of concern is the potential hazard of transfer of contraceptive steroids to the infant (a significant amount of the progesterational component is transferred into breast milk). However, no adverse effects have thus far been identified. Women who use oral contraceptive have a lower incidence of breastfeeding after the 6th postpartum month, regardless of whether oral contraception was started at the first, second, or third postpartum month. This study also found that mothers on birth control pills lactated a significantly shorter period of time than controls, a mean of 3.7 months versus 4.6 months in controls.

Because the above considerations indicate that oral contraception shortens the duration of breastfeeding, it is worthwhile to consider the contraceptive effectiveness of lactation. The contraceptive effectiveness of lactation, i.e., the length of the interval between births, depends on the level of nutrition of the mother (if low, the longer the contraceptive interval), the intensity of suckling, and the extent to which supplemental food is added to the infant diet. 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. With full or nearly full breastfeeding, approximately 70% of women remain amenorrheic through 6 months and only 37% through one year; nevertheless with exclusive breastfeeding, the contraceptive efficacy at one year is high, at 92%. Fully breastfeeding women commonly have some vaginal bleeding or spotting in the first 6 postpartum weeks, but this bleeding is not due to ovulation.

Supplemental feeding increases the chance of ovulation (and pregnancy) even in amenorrheic women. Total protection is achieved by the exclusively breastfeeding woman for a duration of only 10 weeks. Half of women studied who are not fully breastfeeding ovulate before the 6th week, the time of the traditional postpartum visit; a visit during the 3rd postpartum week is strongly recommended for contraceptive counseling.

It is apparent that although lactation provides a contraceptive effect, it is variable and not reliable for every woman. Furthermore, because frequent suckling is required to maintain full milk production, women who use oral contraception and also breastfeed less frequently (e.g., because they work outside their home) have two reasons for decreased milk volume. This combination can make it especially difficult to continue nursing.

Initiation of Oral Contraception in the Postpartum Period

The individual woman is in need of contraception early in the postpartum period. In a careful study of 22 postpartum, nonbreastfeeding women, the mean time from delivery to the first menses was 45 ± 10.1 days, and no woman ovulated before 25 days after delivery. A high proportion of the first cycles (81.8%) and the subsequent cycles (37%) were not normal; however, these were not certain not predictable in individual women. Others have documented a mean delay of 7 weeks before resumption of ovulation, but half of the women studied ovulated before the 6th week, the time of the traditional postpartum visit. The obstetric tradition of scheduling the postpartum visit at 6 weeks should be changed. A 3-week visit would be more productive in avoiding postpartum surprises.

The Rule of 3’s:

In the presence of FULL breastfeeding, a contraceptive method should be used beginning in the 3rd postpartum month.

With PARTIAL breastfeeding or NO breastfeeding, a contraceptive method should begin during the 3rd postpartum week.

After the termination of a pregnancy of less than 12 weeks, oral contraception can be started immediately. After a pregnancy of 12 or more weeks, oral contraception has traditionally been started 2 weeks after delivery to avoid an increased risk of thrombosis during the initial postpartum period. We believe that oral contraception can be started immediately after a second-trimester abortion or premature delivery.

Because of the concerns regarding the impact of oral contraceptives on breastfeeding, a useful alternative is to combine the contraceptive effect of lactation with the progestin-only minipill. This low dose of progestin has no negative impact on breast milk, and some studies document an increase in milk quantity and nutritional quality. Highly effective (near total) protection can be achieved with the combination of lactation and the minipill. Because of the slight positive impact on lactation, the minipill can be started soon after delivery, but at least a 3-day postpartum delay is recommended to allow the decline in pregnancy levels of estrogen and progesterone and the establishment of lactation. In addition, use of the progestin-only minipill has been reported to be associated with a 3-fold increased risk of diabetes mellitus in lactating women with recent gestational diabetes. This special group of women should consider other methods of contraception.

Other Considerations

Prolactin-Secreting Adenomas

Because estrogen is known to stimulate prolactin secretion and to cause hypertension 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. 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. Previous use of oral contraceptives is not related to the size of prolactinomas at presentation and diagnosis. Oral contraception can be prescribed to patients with pituitary microadenomas without fear of subsequent tumor growth. 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, 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. 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. 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 oligoovulation 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. Again, the most important concern is and should be the prevention of an unwanted pregnancy. For most teenagers, oral contraception, dispensed in the 28-day package for better compliance, is the contraceptive method of choice.
Infections and Oral Contraception

Viral STDs

The viral STDs 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 STDs. 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. \textsuperscript{22, 55, 276} For women not in a stable, monogamous relationship, a dual approach is recommended, combining the contraceptive efficacy and protection against PID offered by oral contraception with the use of a barrier method (and spermicide) for prevention of viral STDs.

Bacterial STDs

Sexually transmitted diseases (STDs) are one of the most common public health problems in the United States. It was estimated in 1995, that 7.6% of reproductive age U.S. women reported having been treated for pelvic inflammatory disease (PID). \textsuperscript{22} This upper genital tract infection is usually a consequence of STDs. The best estimate of subsequent tubal infertility is derived from an excellent Swedish report; approximately 12% after one episode of PID, 23% after 2 episodes, and 54% after 3 episodes. \textsuperscript{22} 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. \textsuperscript{55, 221, 276} 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. \textsuperscript{55, 221} Furthermore, if a patient does get a pelvic infection, the severity of the salpingitis found at laparoscopy is decreased. \textsuperscript{55, 221} 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."

The argument has been made that this protection is limited to gonococcal disease, and chlamydial infections may even be enhanced. Fifteen of 17 published studies by 1985 reported a positive association of oral contraceptives with lower genital tract chlamydial cervicitis. \textsuperscript{55} Because lower genital tract infections caused by chlamydia are on the rise (now the most prevalent bacterial STD 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 disease (multiple sexual partners, a history of STD, 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. \textsuperscript{55}

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. \textsuperscript{55} In fact, a case-control study indicated that oral contraceptive users with chlamydia infection are protected against symptomatic PID. \textsuperscript{55} A case-control study has 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. \textsuperscript{55} 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, and evidence indicates that this is so. \textsuperscript{55}

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 6th 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. \textsuperscript{22} Evidence is lacking to convincingly implicate oral contraception with vaginal infections with Candida species; \textsuperscript{22} 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. Known or suspected breast cancer.
4. Undiagnosed abnormal vaginal bleeding.
5. Known or suspected pregnancy.
6. Smokers over the age of 35.

Relative Contraindications Requiring Clinical Judgment and Informed Consent

1. Migraine headaches. In retrospective studies of low-dose pills, it is not clear whether migraine headaches are associated with an increased risk of stroke. Some women report an improvement in their headaches, and in our view, a trial of the lowest dose oral contraceptives is warranted. 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). \textsuperscript{55}
2. Hypertension. A woman under 35 who is otherwise healthy and whose blood pressure is well controlled by medication can elect to use oral contraception. We recommend the use of the lowest estrogen dose products. \textsuperscript{55}
3. Uterine leiomyoma. This is not a contraindication with 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. \textsuperscript{22} 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. \textsuperscript{55, 221, 276} The administration of low-dose oral contraceptives to women with leiomyomatosis does not stimulate fibroid growth, and is associated with a reduction in menstrual bleeding. \textsuperscript{55}
4. 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. \textsuperscript{55, 55} We believe that women with previous gestational diabetes can use oral contraception with annual assessment of the fasting glucose level.
5. 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. It is prudent to maintain contraception right up to the performance of a sterilization procedure, and this short, outpatient operation carries very minimal risk.
6. Epilepsy. Oral contraceptives do not exacerbate epilepsy, and in some women, improvement in seizure control has occurred. \textsuperscript{55, 276} Antiepileptic drugs, however, may decrease the effectiveness of oral contraception.
7. Obstructive jaundice in pregnancy. Not all patients with this history will develop jaundice on oral contraception, especially with the low-dose formulations.
8. Sickle cell disease or sickle C disease. Patients with sickle cell trait can use oral contraception. The risk of thrombosis in women with sickle cell disease or sickle C disease is theoretical (and medicolegal). We believe effective protection against pregnancy in these patients warrants the use of low-dose oral contraception.

Diabetes mellitus. Effective prevention of pregnancy outweighs the small risk of complicating vascular disease in diabetic women who are under age 35 and otherwise healthy.
10. **Gallbladder disease.** Oral contraceptives do not cause gallstones, but may accelerate the emergence of symptoms when gallstones are already present.

**Clinical Decisions**

**Surveillance**

In view of the increased safety of low-dose preparations for healthy young women with no risk factors, such 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. 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.

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 5th day of the cycle, and no pills are missed. Thus, starting oral contraception on the first day of menses ensures immediate protection. In the United States, most clinicians and patients prefer the Sunday start packages, beginning on the first Sunday following menstruation. This 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. Some clinicians prefer to advise patients to use added protection in the first week of use.

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 7-day hormone-free interval. Simply start a new package of pills the next day after finishing the series of 21 pills in the previous package. Remember, when using a 28-pill package, the patient would start a new package after using the 21 active pills.

**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.

**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. 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 is needed.

If she misses 2 pills in the first two weeks, she should take two pills on each of the next two days; it is unlikely that a back-up method is needed, but the official consensus is to recommend backup for the next 7 days.

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. Studies in which women deliberately lengthen their pill-free interval up to 11 days have failed to show signs of ovulation. So far there is no evidence that moving to lower doses has had an impact on the margin of error. The studies have involved small numbers of women and given the large individual variation, it is still possible that some women might be at risk with a small increase in the pill-free interval. However, the progestational effects on endometrium and cervical mucous serve to ensure good contraceptive efficacy. We may well prove that current recommendations are too conservative, and that a woman’s chance of getting pregnant with missing pills is nearly zero. Nevertheless, this conservative advice is the safest message to convey.

The most prevalent problems that can be identified associated with apparent oral contraceptive failures are vomiting and diarrhea. 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 years 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.

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. However, breakthrough bleeding is further increased in women who smoke and use formulations with 20 µg ethinyl estradiol. Breakthrough bleeding is best managed by encouragement and reassurance. This bleeding usually disappears by the third cycle in the majority of women. 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 progesterin-induced decidualization. This endometrium and blood vessels within the endometrium will 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 from beginning to later cycles. Reinforcement of consistent pill taking can help minimize breakthrough 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.

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).

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; a basal body temperature less than 98 degrees (36.6°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. This is obviously a problem of perception. 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. The low progestin doses (including levonorgestrel formulations) currently used are insufficient to stimulate an androgenic response.

**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. Functional ovarian cysts occurred less frequently in women on higher dose oral contraception. This protection is reduced with the current lower dose products to the point where little effect can be measured.

**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. 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).

There is good reason to believe that drugs, which stimulate the liver's metabolic capacity, can affect oral contraceptive efficacy. 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.

To be cautious, patients on medications that affect liver metabolism should choose an alternative contraceptive. These drugs are as follows:

- Rifampin
- Phenobarbital
- Phenytoin (Dilantin)
- Prinidone (Mysoline)
- Carbamazepine (Tegretol)
- Possibly 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. 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.

**Migraine Headaches**

True migraine headaches are more common in women, while tension headaches occur equally in men and women. There have been no well done studies to determine the impact of oral contraception on migraine headaches. Patients may report that their headaches are worse or better.

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. 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.

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. Switching to a different brand is worthwhile, if only to evoke a placebo response. True vascular headaches (classic 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).

**Clues To Severe Vascular Headaches:**

- Headaches that last a long time.
- Dizziness, nausea, or vomiting with headaches.
- Scotomata or blurred vision.
- Episodes of blindness.
- Unilateral, unremitting headaches.
- Headaches that continue despite medication.

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. We have had personal success (anecdotal to be sure) alleviating headaches by eliminating the menstrual cycle, either with the use of daily oral contraceptives or the daily administration of a progesterational agent (such as 10 mg medroxyprogesterone acetate) or the use of depot-medroxyprogesterone acetate. Some women with migraine headaches have extremely gratifying responses. Women who experience an exacerbation of their headaches with oral contraception should consider one of the progestin-only methods.

**Summary: Oral Contraceptive Use and Medical Problems**

**Gestational Diabetes.** There is no contraindication to combined oral contraceptive use following gestational diabetes. 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. In a one-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.

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.

Hemorrhagic Disorders. Women with hemorrhagic disorders and women taking anticoagulants can use oral contraception. 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.

Galbladder 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.

Obesity. An obese woman who is otherwise healthy can use low-dose oral contraception.

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.

Seizure Disorders. There is no impact of oral contraceptives on pattern or frequency of seizures. The concern is that anticonvulsant-induced hepatic enzyme activity can increase the risk of contraceptive failure. Some clinicians advocate the use of higher dose (50 mg estrogen) products; however, no studies have been performed to demonstrate that this higher dose is necessary.

Mitril 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 erythematosus, and the vascular disease associated with lupus, when present, represents a contraindication to estrogen-containing oral contraceptives.\(^{22}\) The progestin-only methods are a good choice. However, in patients with stable or inactive disease, without renal involvement and high antiphospholipid antibodies, low-dose oral contraception can be considered.\(^{22}\) SELENA (Safety of Estrogen in Lupus Erythematosus National Assessment) is an on-going randomized, controlled clinical trial of oral contraceptive therapy in premenopausal women with systemic lupus erythematosus (as well as postmenopausal hormone therapy).

Migraine Headaches. Low-dose oral contraception (the lowest estrogen dose formulation) can be tried with careful surveillance in women with common migraine headaches. Daily administration can prevent menstrual migraine headaches. Oral contraception is best avoided in women with classic migraine headaches associated with neurologic symptoms.

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 medicolegal). 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).\(^{22}\) 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.\(^{113}\)

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.

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). 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.\(^{22}\) Oral contraceptives containing desogestrel, noregestimate, or gestodene can increase HDL levels, but it is not known if this change is clinically significant.

Depression. Low-dose oral contraceptives have minimal, if any, impact on mood.

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 one 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 may be a better choice for smoking women, regardless of age (because this dose of estrogen has no impact on clotting factors and platelet activation).\(^{12}\)

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.

Ulcerative Colitis. There is no association between oral contraception and ulcerative colitis. Women with this problem can use oral contraceptives. Oral contraceptives are absorbed mainly in the small bowel.

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. 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.\(^{12}\) However, a large clinical trial has demonstrated typical contraceptive efficacy with one pill administered vaginally per day.\(^{12}\)

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, progesterin 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
Oral contraceptives are frequently utilized to manage the following problems and disorders:

- Rheumatoid arthritis.
- Osteoporosis.
- Endometriosis.
- Benign breast disease.
- Pelvic inflammatory disease.
- Hormone therapy for hypothalamic amenorrhea.
- Acne and hirsutism.
- Endometriosis prophylaxis.
- Dysmenorrhea.
- Dysfunctional uterine bleeding.
- Acne and hirsutism.
- Hormone therapy for hypothalamic amenorrhea.
- Prevention of menstrual porphyria.
- Control of bleeding (dyscrasias, anovulation).

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.

**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.**
**Fewer Ecstatic Pregnancies.**
**More Regular Menses.**

- Less flow.
- Less dysmenorrhea.
- Less anemia.

**Less Salpingitis.**
**Possibly Less Endometriosis.**
**Possibly Less Benign Breast Disease.**
**Possibly Less Rheumatoid Arthritis.**
**Possibly Protection against Atherosclerosis.**
**Possibly Increased Bone Density.**
**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 STDs) 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. High-dose oral contraceptive use decreased the incidence of benign breast disease diagnosed clinically as well as fibrocystic disease and fibroadenomas diagnosed by biopsy; hopefully, the same impact will become evident with current lower dose formulations. A 40% reduction in ovarian cancer and a 50% reduction in endometrial cancer represent substantial protection. Studies with higher dose formulations documented in long-term users a 31% reduction in uterine leiomyomata and, in current users, a 76% reduction in corpus luteum cysts and a 49% reduction in functional ovarian cysts. The impact of low-dose preparations on these problems remains to be accurately measured and may be less. Case-control studies with low-dose oral contraceptives have found no impact on the risk of uterine fibroids, neither increased nor decreased. Epidemiologic studies have indicated that a progressive decline in the incidence of ovarian cysts is proportional to the steroid doses in oral contraceptives. Current low-dose monophasic and multiphasic formulations provide no protection against functional ovarian cysts. 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. 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. These benefits involving two common gynecologic problems have an important, positive impact on compliance.

An Austrian study concluded that osteoporosis occurs later and is less frequent in women who have used long-term oral contraception. Cross-sectional studies of postmenopausal women 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. However, other studies reflecting modern use of low-dose products indicate little impact of oral contraceptive use on bone. These measurements of bone density are not as important as the clinical outcome: fractures. The available evidence suggests that any favorable effects on bone density are not clinically important. In the Royal College of General Practitioners Study, the overall risk of fractures in ever users of oral contraceptives was actually slightly increased. Similar results have been observed in the Oxford Family Planning Association Study. It is likely that the increased risk reflects lifestyle effects among oral contraceptive users, but thus far, there is no evidence of a protective effect against fractures. However, previous oral contraceptive users are just now becoming elderly and reaching the age of greatest fracture prevalence. Future studies of postmenopausal women may eventually reveal a beneficial effect on 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. 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 the evidence in 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.

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. 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 prophyllaxis against the recurrence of endometriosis in a woman who has already undergone more vigorous treatment with surgery or the GnRH analogues. To protect against endometriosis, oral contraceptives should be taken daily, with no break and no withdrawal bleeding.

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. The beneficial clinical effect is the same with low-dose preparations containing levonorgestrel, previously recognized to cause acne at high dosage. 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. Theoretically, these products would be more effective in the treatment of acne and hirsutism; however, this is yet to be 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 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. No advantage for the contraceptive treatment could be demonstrated. The cysts resolved completely and equally fast in both treated and non-treated 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. 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 (close to 1 million) occur each year in the United States because of the failure of oral contraception. Worldwide, literally millions of unintended pregnancies result from poor compliance. In general, unmarried, poor, and minority women are more likely to have failures, reaching rates of 10–20%. Overall, the failure rate with actual use ranges from 3 to 6%. This difference between the theoretical efficacy and actual use reflects compliance and 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 3 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. Because these complaints respond well even to placebo treatment, it is reasonable to expect a favorable response to sensitive and attentive counseling.
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, and difficulties arising from the patient package insert.

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 nonconceptive 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).
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.
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. Inconsistent use of oral contraceptives is more common in women who are new starters.
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.

**The Progestin-Only Minipill**

The minipill contains a small dose of a progestational agent and must be taken daily, in a continuous fashion. There is no evidence for any difference 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 norgestrel.
3. Ovrette, Neogest --- 0.075 mg levonorgestrel.
4. Exilut 0.500 ------------------------ mg lynestrenol.
5. Femulen 0.500 ------------------------ mg ethynodiol diacetate.
6. CerazetteR ------------------------ 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 on only those tissues very sensitive to the female sex steroids, estrogen and progesterone. The contraceptive effect is more dependent upon 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. Tubal physiology may also be affected, but this is speculative.

**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 hours to take effect, and, most importantly, the impermeability diminishes 22 hours after administration, and by 24 hours sperm penetration is essentially unpaired.

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).[^2] and there is an immediate return to fertility on discontinuation (unlike the delay seen with the combination oral contraceptive). 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).[^1] 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.[^3] 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).[^4] In motivated women, the failure rate is comparable to the rate (less than 1 per 100 woman-years) with combination oral contraception. [^5][^6][^7]

### Pill Taking

The minipill should be started on the first day of menses, and a backup method must be used for the first 7 days because some women (very few) ovulate as early as 7–9 days after the onset of menses. The pill 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 back-up 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 hours late in taking a pill, a backup method should be used for 48 hours.

### 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%), short, irregular cycles (40%), or a total lack of cycles ranging from irregular bleeding to spotting and amenorrhea (20%). This is the major reason why women discontinue the minipill method of contraception.

Women on progestin-only contraception develop more functional, ovarian follicular cysts.[^8] 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). 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 where 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.[^9] 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.[^10] 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.[^11][^12] In fact, there is a modest positive impact; women using the minipill breastfeed longer and add supplementary feeding at a later time.[^13] Because of the slight positive impact on lactation, the minipill can be started soon after delivery, but at least a 3-day postpartum delay is recommended to allow the decline in pregnancy levels of estrogen and progestrone and the establishment of lactation.[^14]

The minipill is a good choice in situations where estrogen is contraindicated, such as patients with serious medical conditions (diabetes with vascular disease, severe systemic lupus erythematosus, 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. The World Health Organization case-control study could find no indication for increased risks of stroke, myocardial infarction, or venous thromboembolism with oral progestin-only contraceptives.[^15] No impact can be measured on the coagulation system.[^16] The minipill can probably be used in women with previous episodes of thrombosis, and the package insert in the United States has been 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 contraception 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:

- Rifampin
- Phenytoin (Dilantin).
- Primidone (Mysoline).
- Phenobarbital.
- Phenytoin (Dilantin).
- Carbamazepine (Tegretol).
- Possibly ethosuximide, griseofulvin, and troglitazone.

Do the noncontraceptive benefits associated with combination oral contraceptives 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), but probably at reduced levels.

Good efficacy with the minipill requires regularly, 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 large doses of estrogen to prevent implantation 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.[^17] It was quickly appreciated that these extremely large doses of estrogen were associated with a high rate of gastrointestinal side effects. Yuzpe developed a method utilizing a combination oral contraceptive, resulting in an important reduction in dosage.[^18] The following treatment regimens have been documented to be effective:
For women with a contraindication to exogenous estrogen, but this dose is equivalent to 10 pills of the levonorgestrel progestin-only minipill. 35 36 In some countries, special packages of 0.75 mg levonorgestrel are available for emergency contraception. Greater efficacy and fewer side effects make low-dose levonorgestrel the treatment of choice. In the United States, a kit is available containing 4 tablets, each containing 50 µg ethinyl estradiol and 0.250 mg levonorgestrel, to be used in the usual fashion, 2 tablets followed by 2 tablets 12 hours later. This method has been more 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. 34 35 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% to 800,000 per year. 36

Many women do not know of this method, and it has been difficult to obtain. 35 37 In Europe and New Zealand, special packages with printed instructions are marketed specifically for emergency contraception, and this is now available in the U.S. Even if women are aware of this method, accurate and detailed knowledge is lacking. 32 A favorable attitude toward this method requires knowledge and availability. Information can be obtained from the following web site maintained by the Office of Population Research at Princeton University: http://opr.princeton.edu/ec.

Clinicians should consider providing emergency contraceptive kits 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 a study of such an approach, self-administration by appropriately screened and educated women was found to be effective and free of unwanted effects. 38

**Mechanism and Efficacy**

The mechanism of action is not known with certainty, but it is believed with justification that this treatment combines delay of ovulation with a local effect in the endometrium. 39 40 The efficacy has been confirmed in large clinical trials and summarized in complete reviews of the literature. 41 42 43 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 hours (0.1%), but the side effects make the combination oral contraceptive a better choice. In general clinical use, the method 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 44) yields the 2% failure rate measured in clinical studies (in other words, 98%, effective). 45 46 Results with levonorgestrel will be even better, 99% effective.

**Treatment Method**

Treatment should be initiated as soon after exposure as possible, and the standard recommendation is that it be no later than 72 hours. 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 might extend beyond 72 hours. 47 Data from randomized, clinical trials, however, support the importance of timing, finding a reduction in efficacy after 72 hours. 48 Because of possible, but unlikely, harmful effects of these high doses to a fetus, an already existing pregnancy should be ruled out prior to use of postcoital hormones. Furthermore, the patient should be offered induced abortion if the method fails. This patient encounter also provides an important opportunity to screen for STDs.

The combination oral contraceptive method delivers significantly less steroid hormone, and this reduction in the total dose and the number of doses reduces the side effects and limits them to a shorter time period. It is worth adding an antiemetic, oral or suppository, to the treatment; the long-acting nonprescription agent, meclizine, is recommended, to be taken one hour before the emergency contraception treatment. Side effects reflect the high doses used: nausea (50%), vomiting (20%), breast tenderness, headache, and dizziness. If a patient vomits within an hour after taking pills, additional pills must be administered as soon as possible. Although short-term treatment with combined oral contraceptives has been documented to have no effect on clotting factors, the usual contraindications for oral contraception apply to this use. 49 In view 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 minipill can be used for emergency contraception; e.g., administering 10 levonorgestrel tablets (75 µg), for each of the two doses, or in some countries using the special commercial package.

A 3-week follow-up visit should be scheduled to assess the result, and to counsel for routine contraception.

Could other combination oral contraceptive products be used? Because other doses and other formulations have never been tested, the efficacy is unknown. It would not be appropriate to expose patients to an unknown failure rate. The use of danazol for emergency contraception is not effective. 50 Mifepristone (RU486), the progesterone antagonist, has been without failures and with lower side effects in clinical trials.

The 3 major problems with the available methods of emergency contraception are the high rate of side effects, the need to start treatment within 72 hours after intercourse, and the small, but important, failure rate. Mifepristone in a single oral dose of 600 mg is associated with markedly less nausea and vomiting and an efficacy rate of nearly 100%. 51 52 Mifepristone is used for emergency contraception in China in a dose as low as 50 mg. Clinical studies have indicated that doses as low as 5 mg or 10 mg daily are effective. 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.

Another method of emergency contraception is the insertion of a copper IUD, up to 5 days after unprotected intercourse. The failure rate (in a small number of studies) is very low, 0.1%. 53 54 This method definitely prevents implantation, but it is not suitable for women who are not candidates for intrauterine contraception, e.g., multiple sexual partners, rape victim.

**Oral Contraception for Older Women**

Women of the post-World War II generation have faced a unique evolutionary change. They were the first to be able to exercise control over their fertility, and then as they aged and deferred pregnancy, they had to deal with the problem of unintended infertility. After World War II, the U.S. total fertility rate reached a modern high of 3.8 births per woman. The last women born in this period will not reach their 45th birthday until around 2010. For approximately a 20-year period, therefore, there will be an unprecedented number of women in the later child-bearing years. The aging of the World War II population boom is giving current times a greater number of women who are delaying marriage and childbirth. This demographic change has 3 specific impacts on couples.

Ovral: 2 tablets followed by 2 tablets 12 hours later.

Alesse: 5 tablets followed by 5 tablets 12 hours later.

Lo Ovral, Nordette, Levien, Triphasil, Trilevlen: 4 tablets followed by 4 tablets 12 hours later.
A need for effective contraception.

2. The problem of achieving pregnancy later in life.
3. The problem of being pregnant later in life.

This combination of increasing numbers, deferment of marriage, and postponement of pregnancy in marriage is responsible for the fact that we will be seeing more and more older women who will need reversible contraception. This is underscored by the fact that from ages 20–44, American women have the highest proportion of pregnancies aborted compared to other countries, indicating an unappreciated, but real, problem of unintended pregnancy existing beyond the teenage years, especially after age 35. More than half of all pregnancies in the U.S. are estimated to be unplanned, and more than half of these are aborted. The best way to minimize the number of induced abortions is effective contraception.

From 1970 to 1986, the number of births in women over 30 quadrupled; however, since 1990, the fertility rate among women over 30 has remained relatively stable. As more and more couples defer pregnancy until later in life, the use of sterilization under age 35 will decline, and the need for reversible contraception will increase. Between 1988 and 1995, oral contraceptive use decreased in women younger than 25 and increased in women aged 30–44. These numbers changed because clinicians and patients have come to understand and accept that low-dose oral contraception is safe for healthy, nonsmoking older women.

Oral Contraception for the Transition Years

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 oral contraception 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 that will last 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. Occasionally, many women will live through a period of anovulation. Corpus luteum formation and function occur, and therefore the older woman is not totally safe from the threat of an unplanned and unexpected pregnancy.

Fortunately clinicians and patients have recognized that low-dose oral contraception is very safe for healthy, nonsmoking older women. However, their use is still not sufficient to meet the need. Besides fulfilling a need, 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.

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. 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 oral 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 consider using 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. In a randomized study of women over age 30, this formulation was associated with the virtual elimination of any effects on coagulation factors. Indeed, the 20 µg formulation has no significant impact on the measurements of clotting factors, even in smokers.

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 probably negligible for healthy nonsmoking 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 women, 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.

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 one week before menses is simple enough to obtain and worth the cost. The prompt diagnosis of anovulation (serum progesterone less than 300 ng/dL) 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 10 mg medroxyprogesterone acetate given daily the first 10 days of each month. If hyperplasia is already present, follow-up aspiration office curettage after 3–4 months is required. If progestin treatment is ineffective and histological 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.

If contraception is desired, the clinician and patient should seriously consider the use of oral 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.

Climinicians have been made so wary of providing oral contraceptives to older women that a traditional postmenopausal hormone regimen is often utilized to treat women with the kind of irregular cycles usually expected in the transitional years. This age 35 will exogenously and the need for a variable of estrogen 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 something that is often unappreciated, the standard doses of estrogen and progestin in a postmenopausal regimen will not suppress gonadotropins and prevent ovulation.

When to Change From Oral Contraception to Postmenopausal Hormone Therapy

A common clinical dilemma is when to change from oral contraception to postmenopausal hormone therapy. It is important to change because even with the lowest estrogen dose oral contraceptive available, the estrogen dose is four-fold greater than the standard postmenopausal dose, and with increasing age, the dose-related
Effects of newer oral contraceptives on the inhibition of coagulation and fibrinolysis in relation to dosage and type of steroid,

High risk of cerebral-vein thrombosis in carriers of a prothrombin-gene mutation and in users of oral contraceptives,

Oral contraceptives and thromboembolic disease: effects of lowering oestrogen content,

Spitzer WO, Lewis MA, Heinemann LAJ, Thorogood M, MacRae KD, on behalf of the Transnational Research Group on Oral Contraceptives and the Health of Young Women,

Gerstman BB, Piper JM, Tomita DK, Ferguson WJ, Stadel BV, Lundin FE,

Rosendaal FR, Doggen CJM, Zivelin A, Arruda VR, Aiach M, Siscovick DS, Hillarp H, Watzke HH, Bernardi F, Cumming AM, Preston FE, Reitsma PH,

Changes in coagulation and anticoagulation in women taking low-dose triphasic oral contraceptives: a controlled comparative 12-month clinical trial,

WHO Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception,

Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral

Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C,

Activated protein C resistance due to a common factor V gene mutation is a major risk factor for venous thrombosis,

Lawson DH, Davidson JF, Jick H,

WHO Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception,

Stanczyk FZ,

Oral contraceptive use in relation to nonfatal myocardial infarction,

Notelovitz M, Kitchens CS, Khan FY,

Rees DC, Cox M, Clegg JB,

Trussell J, Hatcher RA, Cates Jr W, Stewart FH, Kost K,

Colditz GA, and the Nurses’ Health Study Research Group,

Lewis MA, Heinemann LAJ, MacRae KD, Bruppacher R, Spitzer WO,

Lidegaard Ø,

Vandenbroucke JP, Rosendaal FR,

Oral contraceptive estrogen dose and the risk of deep venous thromboembolic disease,

Large increase in plasmatic 11-dehydro-TXB

Winkler UH, Schindler AE, Endrikat J, Düsterberg B,

Child TJ, Rees M, MacKenzie IZ,

Pregnancy terminations after oral contraception scare,

Pill method failures,

Pharmacokinetics of the new progestogens and influence of gestodene and desogestrel on ethinylestradiol metabolism,


Skjæretved FE, Increased number of induced abortions in Norway after media coverage of adverse vascular events from the use of third-generation oral contraceptives, Contracept 55:27, 1997.

Spitzer WO, Rossenfald CR, Reitsma PH, Berkman RM, The importance of genetic variation in the 3¢-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thromboembolic events, Blood 88:3647, 1996.


Skjæretved FE, Increased number of induced abortions in Norway after media coverage of adverse vascular events from the use of third-generation oral contraceptives, Contracept 55:27, 1997.


Poort SR, Rossenfald FR, Reitsma PH, Berkman RM, A single genetic origin for the 3¢-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thromboembolic events, Blood 88:3647, 1996.


Ressequie LJ, Hick JF, Bruen JA, Noller KL, O'Fallon WM, Kurland LT, Transfer of norethisterone (NET) and levonorgestrel (LNG) from a single tablet into the infant's circulation through the mother's milk, Contraception 35:617, 1987.


Ressequie LJ, Hick JF, Bruen JA, Noller KL, O'Fallon WM, Kurland LT, Transfer of norethisterone (NET) and levonorgestrel (LNG) from a single tablet into the infant's circulation through the mother's milk, Contraception 35:617, 1987.


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, the Norplant system and depot-medroxyprogesterone acetate (Depo-Provera). Once-a-month injectable contraceptives are used in many parts of the world, and are a new option for U.S. women.

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. Norplant was first introduced into clinical trials in Chile in 1972. Assessment of this method was completed in more than 45 countries, and in 1990, Norplant was approved for marketing in the U.S., the 20th country of approximately 60 to do so.¹

The progestin in Norplant, circulating at levels one-fourth to one-tenth of those obtained with combined oral contraceptives, prevents conception by suppressing ovulation and thickening cervical mucus to inhibit sperm penetration so that fertilization rarely occurs. ¹ Because serum levels of progestin remain low and because no estrogen is administered, this long-acting contraceptive method does not cause any serious health effects. ¹ This method does, however, cause some bothersome side effects attributable to sustained administration of progestin, such as changes in menstrual pattern, weight gain, headache, and effects on mood.

Injectable medroxyprogesterone acetate (Depo-Provera) is a long-acting (3–6 months) agent that has been part of the contraceptive programs of many countries for more than 20 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.

The long-acting progestin methods (including Norplant and Depo-Provera) are as effective as sterilization and IUDs, and more effective than oral and barrier contraception.² 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, theoretical (lowest expected) effectiveness is very close to the actual or typical (use) effectiveness.

Sustained-release methods require less of the user, but they demand more of the clinician. Norplant involves 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.²

The Norplant System

The Norplant subdermal implant system is a long-acting, low-dose, reversible, progestin-only method of contraception for women. It was developed by the Population Council. The implants are manufactured, under license of the Population Council, by Huhtamaki Oy/Leiras Pharmaceuticals in Finland.

The currently available 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 (polyvinylsiloxane and methylvinyl siloxane copolymer) tubing that is sealed shut with silastic medical adhesive (polydimethylsiloxane). The cavity of the capsule has an inner diameter of 1.57 mm, with an inner length of 30 mm. The 6 capsules contain a total of 216 mg levonorgestrel, which is very stable and has remained unchanged in capsules examined after more than 9 years of use.

The implants come packaged in heat-sealed pouches that have a shelf life of 5 years from the date of manufacture and have an additional 5 to 7 years of effective life once inserted. Storage at room temperature with uncontrolled humidity has not altered their composition or lifespan after 4 years, but, optimally, the implants should be stored in a cool, dry area away from direct sunlight. The implants can be ethylene oxide sterilized, but cannot be sterilized by ionizing radiation, dry heat, or autoclaving.

The components of Norplant are not new. The silastic in the tubing has been used in surgical applications such as prosthetic devices, heart valves, and drainage tubes, since the 1950s, and in the most common method of female sterilization (Fallope Rings) since 1970. The progestin, levonorgestrel, has been widely used in oral contraceptives since the 1960s. The toxicity, teratogenicity, and pharmacology of levonorgestrel have been well studied. What is relatively new is the way the system delivers a sustained level of levonorgestrel for a long time — in ongoing trials, up to 7 years, although currently approved for 5 years.

The implants are removed by the clinician during a minor operation. The surgical removal procedure is simple and brief. Although not required, its use can be associated with the discontinuation method. The Norplant system has been in clinical use since 1985, and more than 1 million Norplant systems have been implanted in approximately 100,000 women worldwide.
**Mechanism of Action**

The release rate of the capsule is determined by its total surface area and the thickness of the capsule wall. The levonorgestrel diffuses through the wall of the tubing 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 orally or injected steroids. Within 24 hours after insertion, plasma concentrations of levonorgestrel range from 0.4 to 0.5 ng/mL, high enough to prevent conception; however, a study of cervical mucus changes indicates that a backup method should be used for 3 days after insertion.

The capsules release approximately 85 µg of levonorgestrel per 24 hours during the first 6–12 months of use. This rate declines gradually to 50 µg daily by 9 months and 30 µg per day for the remaining duration of use. The 85 µg of hormone released by the implants during the first few months of use is about equivalent to the daily dose of levonorgestrel delivered by the progesterin-only, minipill oral contraceptive, and 25–50% of the dose delivered by low-dose combined oral contraceptives.

Mean plasma concentrations below 0.20 ng/mL are associated with increased pregnancy rates. 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 5-year mark, mean levels remain above 0.25 ng/mL.

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 use. The greatest decrease over time occurs in women weighing more than 70 kg (154 pounds), but even for heavy women, the release rate is high enough to prevent pregnancy at least as reliably as oral contraceptives.

Levonorgestrel levels may also be affected by the levels of sex hormone-binding globulin (SHBG). Levonorgestrel has a high affinity for SHBG. In the week after Norplant 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 plasma levonorgestrel concentrations.

The mechanism by which Norplant prevents conception is only partially explained. There are 3 probable modes of action, which are similar to those attributed to the contraceptive effect of the progestin-only minipills.

1. The levonorgestrel 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 are ovulatory. 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.
2. The constant level of levonorgestrel has a marked effect on the cervical mucus. The mucus thickens and decreases in amount, forming a barrier to sperm penetration.
3. The levonorgestrel 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.

**Specific Advantages**

Norplant is a safe, highly effective, continuous method of contraception that requires little user effort or motivation and, unlike long-acting injectable contraception, is rapidly reversible. Because this is a progestin-only method, it can be utilized by women who have contraindications for the use of estrogen-containing oral 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. Norplant is not a cotus-related contraceptive method. The use effectiveness closely approximates the theoretical effectiveness. Norplant is an excellent choice for a breastfeeding woman (there is no effect on breastfeeding) and can be inserted immediately postpartum, but at least a 3-day postpartum delay is recommended to allow the decline in pregnancy levels of estrogen and progesterone and the establishment of lactation.

**General Advantages**

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, or lost diaphragms. For women who are at high risk of medical complications should they become pregnant, these methods present a significant safety advantage. Users should be reassured that Norplant use has not been associated with changes in carbohydrate or lipid metabolism, coagulation, liver or kidney function, or immunoglobulin levels. Since many women wanting Norplant will have had negative experiences with other contraceptives, it is important that the differences between this method and previous methods be explained.

**Norplant is a good method for breastfeeding women.** There are no effects on breast milk quality or quantity, and infants grow normally. Another advantage of Norplant is that it allows women to plan their pregnancies precisely; return of fertility after removal is prompt, in contrast to the 18-month delay in ovulation that can follow Depo-Provera injections. In addition, anemia is less likely in Norplant users.

**Disadvantages**

There are some disadvantages associated with the use of the Norplant system. Norplant causes disruption of bleeding patterns in up to 80% of users, especially during the first year of use, and some women or their partners find these changes unacceptable. Endogenous estrogen is nearly normal, and unlike the combined oral 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%, an incidence that can be best minimized by good training and experience in Norplant insertion. Because the insertion and removal of Norplant require minor surgical procedures, initiation and discontinuation costs will be higher than with oral contraceptives or barrier methods.

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. Norplant is not known to provide protection against sexually transmitted diseases (STDs) such as herpes, human papillomavirus, HIV, gonorrhea, or chlamydia. Users at risk for STDs must consider adding a barrier method to prevent infection.

The cost of implants plus fees for insertion total an amount that seems high to many patients unless they compare it with the total cost of using other methods for up to 5 years. 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.

Sustained-release, progestin-only methods do not increase the risk of developing STDs, but it is not known whether they provide any protection. Women who have multiple partners, or whose partners do, should use condoms as well. 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.
Cultural factors can influence the acceptability of menstrual changes. Hispanic users of Norplant, for example, are very accepting of irregular or prolonged bleeding. Some cultures restrict a woman from participating in religious activity, household activities, or sexual intercourse while menstruating.

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.

Prospective users should be allowed to see and touch implants. Women should 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 can be 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.

**Indications**

The Norplant system is indicated for use by women of reproductive age who are sexually active and desire continuous contraception. Norplant should be considered for women who:

1. Desire spacing of future pregnancies.
2. Desire a highly effective, long-term method of contraception.
3. Experience serious or minor estrogen-related side effects with oral 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 do not desire permanent sterilization.
6. Have a history of anemia with heavy menstrual bleeding.
7. Are considering sterilization, but are not yet ready to undergo surgery.
8. Women with chronic illnesses, whose health will be threatened by pregnancy.

**Absolute Contraindications**

Norplant 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, Norplant 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, Norplant, in our view, is 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.

Norplant is not contraindicated in the following situations, but other methods are probably preferable:

1. Severe acne
2. Severe vascular or migraine headaches.
3. Severe depression.
4. Concomitant use of medications that induce microsomal liver enzymes (phenytoin, phenobarbital, carbamazepine, rifampin). In this case, we do not recommend the use of Norplant because of a likely increased risk of pregnancy due to lower blood levels of levonorgestrel.

**Efficacy**

Norplant is a highly effective method of birth control. In studies conducted in 11 countries, totaling 12,133 woman-years of use, the pregnancy rate was 0.2 pregnancies per 100 woman-years of use. All but one of the pregnancies that occurred during this 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.

The overall pregnancy rate after 2 years of use in 9 countries was 0.2 per 100 woman-years of use. The pregnancy rate achieved in the U.S. trials during the second year of use was higher (2.1 per 100 woman-years). Two factors account for this difference. First, users in the U.S. weighed, on the average, more than study participants in other countries. Clinical trials have demonstrated a direct correlation between weight greater than 70 kg (154 pounds) and an increased risk of pregnancy, but even for heavy women, pregnancy rates are lower than with oral contraception. Second, two different types of silastic tubing were used in the manufacture of Norplant capsules. The first type contained a larger proportion of inert filler and was more dense, while the second type contained less filler and was less dense. Higher pregnancy rates have been observed among women using more dense capsules, and in the U.S. trials, capsules were more often of the more dense variety. The less dense tubing is now the only one used in the manufacture of Norplant and has a 15% higher release rate than denser tubing. The 5-year pregnancy rate with the less dense (soft) tubing is comparable to that of tubal ligation.

<table>
<thead>
<tr>
<th>Year</th>
<th>First Pregnancy Rate</th>
<th>Second Pregnancy Rate</th>
<th>Third Pregnancy Rate</th>
<th>Fourth Pregnancy Rate</th>
<th>Fifth Pregnancy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.9%</td>
<td>0.5%</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

Using the less dense tubing, there are now no weight restrictions for Norplant users, but heavier women (more than 70 kg) may experience slightly higher pregnancy rates in the fourth and fifth 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 beyond the 5-year limit. For slender women the duration of Norplant's efficacy may extend well past the fifth year of use. In some extended trials, no pregnancies have occurred into the 7th year.
<table>
<thead>
<tr>
<th>Method</th>
<th>Percent of Women with Pregnancy</th>
<th>Lowest Expected Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No method</td>
<td>85.0%</td>
<td>85.0%</td>
</tr>
<tr>
<td>Combination Pill</td>
<td>0.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Progestin only</td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>IUDs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progesterone IUD</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Levonorgestrel IUD</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Copper T 380A</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Norplant</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Female sterilization</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Male sterilization</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td>Depo-Provera</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Spermicides</td>
<td>6.0</td>
<td>26.0</td>
</tr>
<tr>
<td>Periodic abstinence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calendar</td>
<td>9.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Ovulation method</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Symptothermal</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Post–ovulation</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Withdrawal</td>
<td>4.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Cervical cap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parous women</td>
<td>26.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Nulliparous women</td>
<td>9.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Sponge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parous women</td>
<td>9.0</td>
<td>28.0</td>
</tr>
<tr>
<td>Nulliparous women</td>
<td>6.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Diaphragm and spermicides</td>
<td>6.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Condom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Female</td>
<td>5.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Ectopic Pregnancy

The ectopic pregnancy rate during Norplant use has been 0.28 per 1000 woman-years. 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.

Ectopic Pregnancy Rates per 1000 Woman-Years

<table>
<thead>
<tr>
<th>Ectopic Pregnancy Rates per 1000 Woman-Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-contraceptive users, all ages</td>
</tr>
<tr>
<td>Copper T-380 IUD</td>
</tr>
<tr>
<td>Norplant</td>
</tr>
</tbody>
</table>

Menstrual Effects

Menstrual bleeding patterns are highly variable among users of Norplant. 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%. 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. Studies of the endometrium in Norplant users experiencing abnormal bleeding indicate the presence of enlarged venous sinuses (fragile vessels) and a reduction in the expression of a protein factor (perivascular stromal cell tissue factor) involved in the initiation of hemostasis.

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.

Patients who can no longer tolerate the presence of 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. A therapeutic dose of one of the prostaglandin inhibitors given during the bleeding will help to diminish flow, but estrogen is the
most effective. Another approach is to administer an estrogen-progestin oral contraceptive for 1–3 months.

Although the Norplant system is 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 Norplant are unlikely to become pregnant. 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 dose of levonorgestrel delivered by the implants is not associated with significant metabolic changes. Studies of carbohydrate metabolism, liver function, blood coagulation, immunoglobulin levels, serum cytokines, and blood chemistries have failed to detect changes outside of normal ranges.

No major impact on the lipoprotein profile can be demonstrated. 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 atherosclerosis, just as prolonged exposure to combined oral contraception has not (see Chapter 22). There are no clinically important effects on carbohydrate metabolism. No effect on insulin sensitivity can be detected.

Measurements of bone density in adolescents reveal that Norplant does not affect the teenage gain in bone; similar gains were recorded in Norplant users and control subjects. In older women, an increase in forearm bone density was documented after 6 months of Norplant use.

Effects on Future Fertility

Circulating levels of levonorgestrel become too low to measure within 48 hours after removal of Norplant. 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. The return of fertility after Norplant 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 Norplant and Depo-Provera in the timing of the return to fertility can be critical. Norplant allows precise timing of pregnancy because the return of ovulation after Norplant removal is so prompt. Depo-Provera, 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 Norplant users will have ovulated, but 10 months must elapse before half of Depo-Provera users are ovulatory.

Side Effects

The occurrence of serious side effects is very rare, no different in incidence than observed in the general population. In addition to the menstrual changes, the following side effects have been reported: headache, acne, weight change, mastalgia, hyperpigmentation over the implants, hirsutism, depression, mood changes, anxiety, nervousness, ovarian cyst formation, and galactorrhea.

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. 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. Many complaints respond to reassurance; others can be treated with simple therapies. The most common side effect experienced by users is headache; approximately 20% of women who discontinue use do so because of headache.

Stroke, thrombotic thrombocytopenic purpura, thrombocytopenia, and pseudotumor cerebri have been reported with Norplant. 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 8 countries, no significant excess of cardiovascular events or malignant disease has been observed.

Weight Change

Women using Norplant more frequently complain of weight gain than of weight loss, but findings are variable. In the Dominican Republic, 75% of those who changed weight lost, while in San Francisco, two-thirds gained. 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 should include dietary review and focus on 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).

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 Norplant use. Most Norplant users respond to treatment and do not elect to remove the implants. Careful assessments of the relationship between methylxanthines and mastalgia have failed to demonstrate a link. The most effective treatments 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 Norplant 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 should 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 Norplant 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). This is in contrast to combined oral contraceptives that contain levonorgestrel, where the estrogen effect on SHBG has failed to detect changes outside of normal ranges. However, some continue to grow and cause pain or be palpated at the time of pelvic examination. 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 one month of detection), they need not be sonographically or laparoscopically evaluated. Further evaluation is indicated if they became large and painful or failed to regress. Regular ovulators are less likely to develop ovarian cysts.
likely to form cysts so the situation is likely to improve after two years of Norplant use.

**Herpes Simplex**

Some users have complained of outbreaks of genital herpes simplex lesions occurring more frequently than prior to insertion. Most commonly, the lesions develop during periods of prolonged spotting or bleeding with the wearing of sanitary napkins. Use of vaginal tampons for bleeding and suppression of the virus with oral acyclovir (200 mg tid for up to 6 months) have been successful in dealing with this problem.

**Cancer**

Levonorgestrel and stilastan have been thoroughly evaluated in animals and humans for their carcinogenic effects, and none has been found. Epidemiologic evaluation awaits long-term use by large numbers of women. We can speculate on possible effects of Norplant based on our experience with oral contraceptives and Depo-Provera. 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. The risk of ovarian cancer is also probably reduced, but not as much as with methods that completely suppress ovulation. Breast and cervical cancer effects will be as difficult to assess because of confounding variables as they are with oral contraception and Depo-Provera. The low dose of Norplant, however, would be unlikely to have effects different from other hormonal contraceptives.

**Patient Evaluation**

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 Norplant, 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 should read and sign a written consent for the surgical placement of Norplant. It should include a review of the potential complications of the procedure which 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 should be performed. Based on cervical mucus changes, a backup method need be used no more than 3 days after insertion. Norplant can be inserted immediately postpartum (after 3 days in breastfeeding women), but certainly should be initiated no later than the third postpartum week in non-breastfeeding women and the 3rd postpartum month in breastfeeding women. Acne and headache are less common in women who receive Norplant immediately postpartum, and there is no difference in post-pregnancy weight loss compared with women who receive it 4–6 weeks later.

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.

Selection of the site for placement of Norplant 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 usually 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 immigration of the implants from this site has not been documented. The site of placement does not affect circulating levonorgestrel levels.

**Reasons for Termination**

Although Norplant is a 5-year method, only approximately 30% of women continue use for that long (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. Bothersome side effects, such as menstrual changes, headache, or weight change, are the primary reasons for termination of implant use. Users who cannot tolerate these symptoms request removal in the first two years of use while women who want another pregnancy, the most common personal reason for removal, are more likely to terminate use in the third or fourth year.

Menstrual changes are the most common cause for discontinuation of Norplant in the first year of use. An unsolved concern for many patients and their partners is the fact that bleeding irregularity interferes with sexual interactions. The next most common reasons for discontinuing Norplant are headache, weight change, mood change, anxiety or nervousness, depression, ovarian cyst formation, and lower abdominal pain. Presence of ovarian cysts is not usually an indication for removal of Norplant and is almost never a reason for surgery. Most will resolve spontaneously within 4 weeks. Skin conditions, including rashes, dermatitis, and acne, account for about 0.8% of terminations.

**User Acceptance of Norplant**

Overall, interview surveys throughout the world have indicated that women perceive sustained-release methods, Norplant in particular, as highly acceptable methods of contraception. The most popular feature of Norplant is the ease of use. Approximately 20% of U.S. patients report 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.

In the U.S., the primary motivations for Norplant use have been problems with previous contraceptive methods and ease of Norplant use. Although fear of pain during Norplant 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. Teenagers provide an example of well-documented success. Their one-year pregnancy rates are much lower, and continuation rates much higher than that with oral contraceptives. However, teenage discontinuation of the method due to side effects (especially irregular bleeding and weight gain) is more common with Norplant.

Studies of women's attitudes toward sustained-release contraceptives indicate that the great majority of women find them highly acceptable and perceive them as desirable alternatives to conventional contraceptives, although not everyone finds Norplant preferable. Around the world, women who have used Norplant say they have recommended Norplant to friends and would like to use it again.

**Counseling Women**

Frank information about negative factors such as irregular bleeding and possible weight changes 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 Norplant, for example, before they invest too much time and money in this long-acting contraceptive, is a very important objective of good counseling.

Common patient questions regarding Norplant 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?

LNG ROD — The Two Rod System
The two-implant system, “Norplant-2,” uses levonorgestrel suspended in a silastic matrix and covered with a silastic membrane. Bleeding, pregnancy, and continuation rates were like those of the 6-capsule system, but insertion and removal were easier. However, the manufacturer of the elastomer used in Norplant-2 ceased production. A new system is composed of 2 levonorgestrel rods, called LNG ROD, encased in silicone rubber tubing, sealed at each end with silastic adhesive. Each rod measures 2.5 mm in diameter and 4.3 cm in length, contains 75 mg levonorgestrel, and releases the drug at the same dose as Norplant. A 3-year clinical trial indicates that the performance and side effects are similar to Norplant, but removal is faster and easier.

New Developments in Implant Contraception
The newer prostogens (desogestrel, gestodene, nestorone, nomegestrol, and norgestimate) are less androgenic than levonorgestrel and are useful in contraceptive implants. An example is Implanon, a single implant 4 cm long, that contains 60 mg of 3-keto desogestrel in a core of ethinyl vinyl acetate wrapped with a membrane of the same material. The hormone is released at a rate of about 60 µg per day. Implanon is designed to provide contraception for 2–3 years after which the implant should be removed. Efficacy and side effects are similar to those with Norplant.

Depo-Provera (Depot-Medroxyprogesterone Acetate)
Depo-Provera (depot-medroxyprogesterone acetate) is the most thoroughly studied progestin-only contraceptive. Although its approval for contraception in the U.S. is 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 Depo-Provera 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.

Experience is still too limited to report pregnancy rates, but the longer implant should provide contraception comparable to Norplant.

Biodegradable implants deliver sustained levels of progestin for variable periods of time from a vehicle that dissolves in body tissues. The utility of implant contraception would be improved by the elimination of the need for surgical removal. Two types are under evaluation: Capronor and norethindrone pellets.

Capronor is a single capsule, biodegradable, levonorgestrel-releasing subdermal implant composed of the polymer E-caprolactone. Implants measure 0.24 cm in diameter and either 2.5 or 4 cm in length, providing contraception for one year. The shorter capsule contains 16 mg levonorgestrel, and the longer contains 26 mg. Levonorgestrel escapes from caprolactone at a rate 10 times faster than from silastic. The shorter implant maintains circulating levels of levonorgestrel of 0.2–0.3 ng/mL, while the longer implant maintains higher levels equivalent to those found in Norplant users. The longer capsule suppresses ovulation in a higher proportion (about 50%) of cycles than reported with Norplant, but the shorter implant allows ovulation in most users. The higher release rate allows the use of a smaller implant. Experience is still too limited to report pregnancy rates, but the longer implant should provide contraception comparable to Norplant.

When exposed to tissue fluids, E-caprolactone slowly breaks down into E-hydroxycaproic acid, and then finally to carbon dioxide and water. The capsule remains intact during the first 12 months of use, allowing easy removal. After 12 months, the capsule begins to disappear.

Capronor shares the advantages of Norplant with convenience of use and few metabolic effects. There is no adverse impact on the lipoprotein profile. Removal is easier and quicker. The disadvantages are also similar to Norplant: changes in menstrual patterns and the other side effects typical of low-dose, continuous progestin systems. Biodegradable implants could continue to release small, noncontraceptive amounts of hormone after their period of use as a contraceptive has expired.

New Developments in Implant Contraception
1. Each rod measures 2.5 mm in diameter and 4.3 cm in length, contains 75 mg levonorgestrel, and releases the drug at the same dose as Norplant. 
2. A 3-year clinical trial indicates that the performance and side effects are similar to Norplant, but removal is faster and easier.

Another single implant contraceptive is Uniplant, containing 36 mg norethindrone acetate in a 4 cm silastic tube with a 100 µg per day release rate. It provides contraception for one year.

Biodegradable implants deliver sustained levels of progestin for variable periods of time from a vehicle that dissolves in body tissues. The utility of implant contraception would be improved by the elimination of the need for surgical removal. Two types are under evaluation: Capronor and norethindrone pellets.

Capronor is a single capsule, biodegradable, levonorgestrel-releasing subdermal implant composed of the polymer E-caprolactone. Implants measure 0.24 cm in diameter and either 2.5 or 4 cm in length, providing contraception for one year. The shorter capsule contains 16 mg levonorgestrel, and the longer contains 26 mg. Levonorgestrel escapes from caprolactone at a rate 10 times faster than from silastic. The shorter implant maintains circulating levels of levonorgestrel of 0.2–0.3 ng/mL, while the longer implant maintains higher levels equivalent to those found in Norplant users. The longer capsule suppresses ovulation in a higher proportion (about 50%) of cycles than reported with Norplant, but the shorter implant allows ovulation in most users. The higher release rate allows the use of a smaller implant. Experience is still too limited to report pregnancy rates, but the longer implant should provide contraception comparable to Norplant.

When exposed to tissue fluids, E-caprolactone slowly breaks down into E-hydroxycaproic acid, and then finally to carbon dioxide and water. The capsule remains intact during the first 12 months of use, allowing easy removal. After 12 months, the capsule begins to disappear.

Capronor shares the advantages of Norplant with convenience of use and few metabolic effects. There is no adverse impact on the lipoprotein profile. Removal is easier and quicker. The disadvantages are also similar to Norplant: changes in menstrual patterns and the other side effects typical of low-dose, continuous progestin systems. Biodegradable implants could continue to release small, noncontraceptive amounts of hormone after their period of use as a contraceptive has expired.

Although it is unlikely that such low serum levels of progestin would be harmful to users or to their pregnancies, this question needs to be resolved. The degrading implants can be removed in the event of pregnancy.

Anuuelle is a biodegradable norethindrone pellet that, when injected subdermally, is expected to maintain circulating concentrations of this prostogen at contraceptive levels for up to 3 years. The pellets are composed of 10% pure cholesterol and 90% norethindrone, and are about the size of a grain of rice. This method is currently under development to determine the correct size and number of pellets and the cholesterol:hormone ratio necessary to obtain the release rates that provide contraception. Preliminary trials of 2, 3, and 4 pellets have demonstrated that bleeding patterns are disrupted during the first few months of use, then return to normal patterns. Users of 4 and 5 pellets are more likely to be amenorrheic and anovulatory.

Depo-Provera (Depot-Medroxyprogesterone Acetate)
Depo-Provera (depot-medroxyprogesterone acetate) is the most thoroughly studied progestin-only contraceptive. Although its approval for contraception in the U.S. is 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 Depo-Provera 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.

Depo-Provera comes 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. 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.

Depo-Provera 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 prostogen in a sustained-release system like Norplant and a depot system like Depo-Provera is illustrated in the diagram.

Mechanism of Action
The mechanism of action of Depo-Provera 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 prostogen is high enough to effectively block the LH surge, and, therefore, ovulation does not occur. Suppression of FSH is not as intense as 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. Symptoms of estrogen deficiency, such as vaginal atrophy or a decrease in breast size, do not occur.

The injection should be given within the first 7 days of the current menstrual cycle, otherwise a back-up method is necessary for 1 week. The duration of action can be shortened if attention is not paid to proper administration. The injection must be given deeply in muscle by the Z-track technique and not massaged. It is prudent to avoid locations at risk for massage by daily activities.

Efficacy
The efficacy of this method is equal to that of sterilization, and better than that of all the other temporary methods. Because serum concentrations are relatively high, efficacy is not influenced by weight or by the use of medications that stimulate hepatic enzymes. On the contrary, Depo-Provera is an excellent contraceptive choice for women taking antiepileptic drugs because the high prostogen levels raise the seizure threshold.

Indications
1. At least one year of birth spacing desired.
2. Highly effective long-acting contraception not linked to coitus.
3. Estrogen-free contraception needed.
5. Sickle cell disease.
7. Private, coitally independent method desired.

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 other sustained-release forms of contraception, Depo-Provera 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. Depo-Provera 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 Depo-Provera 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. 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 Depo-Provera users, and a World Health Organization case-control study could find no evidence for increased risks of stroke, myocardial infarction, or venous thromboembolism.

A further advantage exists for patients with sickle cell disease because evidence indicates an inhibition of in vivo sickling with hematologic improvement during treatment. Both the frequency and the intensity of painful sickle cell crises are reduced.

Another advantage is the finding that Depo-Provera increases the quantity of milk in nursing mothers, a direct contrast to the effect seen with combination oral contraception. The concentration of the drug in the breast milk may be negligible, and no effects of the drug on infant growth and development have been observed.

Other benefits associated with Depo-Provera use include a decreased risk of endometrial cancer comparable with that observed with oral contraceptives, and probably the same benefits associated with the progestin impact of oral contraceptives: reduced menstrual flow and anemia, less PID, less endometriosis, fewer uterine fibroids, 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 Depo-Provera users.

Depo-Provera, like oral contraception, may reduce the risk of pelvic inflammatory disease; however, the only study was hampered by small numbers. Suppression of ovulation means that ectopic pregnancies are abolished and ovarian cysts are rare.

The greater the number of choices women have, the more likely they are to find a contraceptive that works well for them. For some women the primary advantages of Depo-Provera 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 don't mind injections. In some societies, injections are respected as efficacious; in these situations, Depo-Provera 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 and intrauterine and implant contraception.
4. Free from estrogen-related problems.
5. Private, use not detectable.
7. Has noncontraceptive benefits.

Problems With Depo-Provera

Major problems with Depo-Provera are irregular menstrual bleeding, breast tenderness, weight gain, and depression. 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. The bleeding is rarely heavy; in fact, hemoglobin values rise in Depo-Provera users. The incidence of irregular bleeding is 70% in the first year, and 10% thereafter. Bleeding and spotting decrease progressively with each re-injection so that after 5 years, 80% of users are amenorrheic (compared with 10% of Norplant users).

Irregular bleeding can be disturbing and annoying, and for many patients, it inhibits sexuality.
Metabolic Effects

The impact of Depo-Provera on the lipoprotein profile is uncertain. While 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. 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. 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 Depo-Provera 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.

Concern has been raised because the blood levels of estrogen with this method of contraception are relatively lower over a period of time compared with a normal menstrual cycle, and, therefore, patients might lose bone to some degree, and, indeed, bone loss has been documented in cross-sectional studies. This bone loss has also been observed in women receiving a high oral dose of medroxyprogesterone acetate, 50 mg daily. However, bone density measurements in women who stopped using Depo-Provera indicated that the loss was regained even after long-term use. Furthermore, a cross-sectional study in Thailand found no bone

For the second time.

For the third time.

For the fourth time.

For the fifth time.
loss in long-term (greater than 3 years) users of Depo-Provera. And most importantly, longitudinal, prospective studies of bone fail to document bone loss in users of Depo-Provera. Loss of forearm bone density could not be detected despite 3 years of Depo-Provera use, suggesting that previous adverse findings could be explained by inadequate control of factors that affect bone, such as smoking and alcohol intake. A small prospective study documented stable forearm bone density over a 6-month period of time. This concern will require ongoing surveillance of past users, but at the present time, this should not be a reason to avoid this method of contraception. It is unlikely that bone loss occurs sufficiently to raise the risk of osteoporosis later in life.

Effect on Future Fertility

The delay in becoming pregnant after ceasing use of Depo-Provera is a problem unique to injectable contraception; all the other temporary methods allow a more prompt return to fertility. However, medroxyprogesterone acetate does not permanently suppress ovarian function, and the concern that infertility with suppressed menstrual function may be caused by Depo-Provera has not been supported by epidemiologic data. The pregnancy rate in women discontinuing the injections because of a desire to become pregnant is normal. By 18 months after the last injection, 90% of Depo-Provera users have become pregnant, the same proportion as for other methods. 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 Depo-Provera. Suppressed menstrual function persists beyond 12 months after the last injection is not due to the drug and deserves evaluation.

Accidental pregnancies occurring at the time of the initial injection of Depo-Provera have been reported to be associated with higher neonatal and infant mortality rates, probably due to an increased risk of intrauterine growth retardation. 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 back-up method is necessary for 2 weeks.

Short-Term Injectable Contraceptives

Monthly or every other month injectable combinations of estrogen and progestin are not new, having been developed over several decades. 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.

Cyclo-Provera (or Cyclofem)

Cyclo-Provera consists of 25 mg depo-medroxyprogesterone acetate and 5 mg estradiol cypionate and is administered monthly as a deep intramuscular injection. This method is as effective as Depo-Provera, but avoids the problems of menstrual irregularity and heavy bleeding, as well as amenorrhea. In addition, the method is rapidly reversible. 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 utilization of an automatic device for self-administration. Approximately 80% of women who are amenorrheic on Depo-Provera will develop vaginal bleeding if switched to Cyclo-Provera.

Norethindrone Enanthate

Norethindrone enanthate is given in a dose of 200 mg intramuscularly every 2 months. This progestin acts in the same way as Depo-Provera, and has the same problems. A combination (Mesigyna) of norethindrone enanthate (50 mg) with estradiol valerate (5 mg) given monthly provides effective contraception with good cycle control. Compared with Cyclofem (25 mg medroxyprogesterone acetate and 5 mg estradiol cypionate), Mesigyna has less bleeding problems. Fertility returns rapidly (by one month) after discontinuation with once-a-month methods.

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 Mesigyna and Cyclofem, the monthly regimen allows regular, and even reduced, cyclic bleeding. 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.

New Developments in Injectable Contraception

Microspheres or microcapsules have been studied for several years. They consist of a biodegradable copolymer and one or more hormones. Like other injectable contraceptives, they are easy to administer and are highly effective. Unlike implants, injectables do not require surgical skills of the clinician and can be discontinued by the patient simply by declining to have another injection. Unlike implants, the microspheres cannot be removed once they are injected. If a woman experiences side effects or becomes pregnant, the hormone will remain in her body until completely metabolized. For this reason, the duration of action of the norethindrone capsules has been limited to a few months.

The microspheres are delivered to the uterus by diffusion and later by degradation of the carrier. The size of the microspheres, the amount of hormone contained within the carrier, and the quantity of microspheres delivered by injection determine the daily dose of norethindrone delivered. Injections currently under evaluation contain a total dose of either 65 mg or 100 mg norethindrone, and the amount released daily is approximately the same as that delivered by low-dose oral contraception, but circulating levels are more stable.

As with other progestin-only methods of contraception, menstrual changes occur and are the most common cause of discontinuation during the first year of use. Users may experience amenorrhea or persistent or irregular spotting or bleeding. In contrast to Depo-Provera, hormone levels decline rapidly after the microspheres have degraded, so that contraceptive effectiveness ends promptly at the predicted time. Most users will resume ovulatory cycles within 2–3 months after the predicted duration of the injection. If pregnancy occurs shortly after expiration of the norethindrone microspheres, the fetus will not be exposed to significant levels of norethindrone.

Microsphere preparations containing norethindrone combined with ethinyl estradiol are also under development. It is hoped that the addition of estrogen at a low dose
Long-Acting Methods for Older Women

The long-acting methods of hormonal contraception deserve consideration in those situations where combination estrogen-progesterin is unacceptable because of health problems or where progesterin has already proved to be unsuitable. These methods are especially advantageous for smokers and for women with a history of thromboembolic disease. Progestogen-only contraception is a good choice for women with hyperlipidemia, for diabetic women (even if they are older and smoke), and for women with severe migraine headaches or hypertension.

Older women, as they approach the menopause, may be more comfortable with the irregular bleeding or amenorrhea associated with these methods. However, the irregular bleeding patterns can cause more concern in some women regarding possible pathology. Postmenopausal hormone treatment can be initiated if menopausal symptoms develop or when annual measurement of the FSH level, beginning at age 50, indicates a rise above 20 IU/L (progestins do not suppress FSH secretion, and, therefore, this measurement can be obtained even while the patient has therapeutic circulating levels of progesterone.

Chapter References


143. Coutinho EM, Spinola P, Barbosa I, Gatto M, Tomaz G, Morais K, Yazile ME, de Souza RN, Neto JSP, de Barros Leal W, Leal C, Hippolito SB, Abranches AD. Multicenter, double-blind, comparative clinical study on the efficacy and acceptability of a monthly injectable contraceptive combination of 150 mg dihydroxyprogesterone acetophenide and 10 mg estradiol enanthate compared to a monthly injectable contraceptive combination of 90 mg dihydroxyprogesterone acetophenide and 6 mg estradiol enanthate. Contraception 55:175, 1997.


Copyright ©
The Intrauterine Device (IUD)

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 which were attached to stems extending into the cervical canal. 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 having a nickel and bronze wire protruding through the cervix. Shortly after, Pust combined Richter’s ring with the old button-type pessary, and replaced the wire with a catgut thread. 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.

The Gräfenberg ring was short-lived, falling victim to Nazi political philosophy that was bitterly opposed to contraception. The non-Aryan Gräfenberg was finally sent to jail, but 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. Ota also fell victim to World War II politics, being 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 the World War II period. 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. 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 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, allowing 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 Margulies Coil was rapidly replaced by the Lippes Loop, which quickly became the most widely prescribed IUD in the United States in the 1970s.

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 being used.

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. The multafilamented 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 ever since, media and the public in the U.S. have 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 following these hearings turned to IUDs, particularly the Dalkon Shield, which was promoted as suitable for multiparous 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 diseases (STDs) and pelvic inflammatory disease (PID) for which IUDs were held partially responsible.

The first epidemiologic studies of the relationship between IUDs and PID used as controls women who depended on oral contraception or barrier methods, and who were, therefore, at reduced risk of PID compared with non-contraceptors and IUD users. 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 of first intercourse, and increased frequency of intercourse). The Dalkon Shield magnified the risk attributed to IUDs because its high failure rate in young women who were already at risk of STDs led to septic spontaneous abortions and, in some cases, death. 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 post-marketing surveillance.

The 1980s saw the decline of IUD use in the United States 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 number of reproductive age women using the intrauterine device decreased by two-thirds from 1982 to 1988, and further decreased in 1995, from 7.1% to 2% to 0.8%, respectively. Nevertheless, in the rest of the world, the IUD is the most widely used method of reversible contraception; currently, over 100 million women use the IUD.

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. Howard Tatum 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 mm² of exposed copper wire), also known as the Tatum-T. 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 mm² of exposed copper surface area) with copper wound around the stem plus a copper sleeve on each horizontal arm. 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.”

The “Progestasert” was developed by the Alza Corporation at the same time that the copper IUDs were developed. This T-shaped device releases 65 µg of progesterone per day for at least one year. The progesterone diminishes the amount of cramping and the amount of blood loss; thus, it is especially useful for women who have heavy periods and cramping. The short lifespan can be and has been solved by using a more potent progestin, such as levonorgestrel.

Efforts continue to develop IUDs that address the main problems of bleeding and cramping. The IUDs of the future will possibly be medicated and frameless.

Types of IUDs

Unmedicated IUDs

The Lippe’s Loop, made of plastic (polyethylene) impregnated with barium sulfate, is still used throughout the world (except in the U.S.). Flexible stainless steel rings
Copper IUDs

The first copper IUDs were wound with 200 to 250 mm² 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 modern generation of IUDs in China includes a stainless steel ring with copper wire that also releases indomethacin (very effective with a low expulsion rate and less blood loss), a V-shaped copper IUD, and a copper IUD shaped like the uterine cavity. The SoT is a copper IUD used only in Switzerland.

The TCu-380A is a T-shaped device with a polyethylene frame holding 380 mm² of exposed surface area of copper. 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 is made of 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.

The Multiload-375 has 375 mm² 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.

The Nova T is similar to the TCu-200, containing 200 mm² 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 WHO data; however, results from Finland and Scandinavia indicate low and stable pregnancy rates over 5 years of use.

Hormone-Releasing IUDs

The Progestasert is a T-shaped IUD made of ethylene/vinyl acetate copolymer containing titanium dioxide. The vertical stem contains a reservoir of 38 mg progesterone together with barium sulfate dispersed in silicone fluid. The horizontal arms are solid and made of the same copolymer. Two blue-black, monofilament strings are attached at a hole in the base of the stem. Progestosterone is released at a rate of 65 µg per day.

The LNG-20, manufactured by Leiras in Finland, releases in vitro 20 µg of levonorgestrel per day. This T-shaped device has a collar attached to the vertical arm, which contains 52 mg levonorgestrel dispersed in polydimethylsiloxane and released at a rate of 15 µg per day in vivo. The levonorgestrel IUD lasts up to 10 years and reduces menstrual blood loss and pelvic infection rates.

Future IUDs

Modifications of the copper IUD are being studied throughout the world. The Ombrette-250 and Ombrette-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 GynFix), consists of 6 copper sleeves (330 mm² 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.

Mechanism of Action

The contraceptive action of all IUDs is mainly in the uterine cavity. Ovulation is not affected, nor is the IUD an abortifacient. 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. If this action should fail, the inflammatory response would also prevent implantation. In women using copper IUDs, sensitive assays for human chorionic gonadotropin (HCG) do not find evidence of fertilization. This is consistent with the fact that the copper IUD protects against both intrauterine and ectopic pregnancies (see below).

The copper IUD releases free copper and copper salts that both have a biochemical and morphological 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 enhanced inflammatory response, marked by production in the endometrium of cytokine peptides known to be cytotoxic. An additional spermicidal effect probably takes place in the cervical mucus.

Efficacy of IUDs

Intrauterine Pregnancy

The TCu-380A is approved for use in the United States for 10 years. However, the TCu-380A has been demonstrated to maintain its efficacy over at least 12 years of use. The TCu-200 is approved for 4 years and the Nova T for 5 years. The progesterone-releasing IUD must be replaced every year because the reservoir of progesterone is depleted in 12–18 months. The levonorgestrel IUD can be used for at least 7 years, and probably 10. The progesterone IUD has a slightly higher failure rate, but the levonorgestrel device that releases 15–20 mg levonorgestrel per day is as effective as the new copper IUDs. 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 copper IUDs.

First Year Clinical Trial Experience in Parous Women
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.

### Ten-Year Experience with Paragard, TCu-380A

<table>
<thead>
<tr>
<th>Device</th>
<th>Rate per 100 users per year</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>0.7</td>
<td>2</td>
</tr>
<tr>
<td>Expulsion</td>
<td>5.7</td>
<td>3</td>
</tr>
<tr>
<td>Bleeding/pain removal</td>
<td>11.9</td>
<td>4</td>
</tr>
<tr>
<td>Medical removal</td>
<td>2.5</td>
<td>5</td>
</tr>
<tr>
<td>Continuation</td>
<td>76.8</td>
<td>6</td>
</tr>
<tr>
<td>Number starting each year</td>
<td>4,923,1492,0181,121,672,621</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>563,483,423,325</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Data from Population Council (n = 3,536) and WHO (n = 1,396) 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 one per 100 women per year. 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). In developing countries, the failure rate with IUDs is less than that with oral contraception. 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. 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).

### Ectopic Pregnancy

The previous use of an IUD does not increase the risk of a subsequent ectopic pregnancy. The current use of an IUD, other than the progesterone-releasing device, offers some protection against ectopic pregnancy. 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. 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). The rate is about one-tenth the ectopic pregnancy rate associated with the Lippes Loop or with devices with less copper such as the TCu-200. The progesterone-releasing IUD has a higher rate, probably because its action is limited to a local effect on the endometrium, while very few ectopic pregnancies have been reported with the levonorgestrel IUD, 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.

The risk of ectopic pregnancy does not increase with increasing duration of use with the TCu-380A or the levonorgestrel IUD. In a 7-year prospective study, not a single ectopic pregnancy was encountered with the levonorgestrel IUD. In 8000 woman-years of experience in randomized multicenter trials, there has been only a single ectopic pregnancy reported with the TCu-380A (which is one-tenth the rate with the Lippes Loop or TCu-200).

The protection against ectopic pregnancy provided by the TCu-380A and the levonorgestrel IUD makes these IUDs acceptable choices for contraception in women with previous ectopic pregnancies.

### Ectopic Pregnancy Rates per 1000 Woman-Years

<table>
<thead>
<tr>
<th></th>
<th>Rate per 1000 Woman-Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-contraceptive users, all ages</td>
<td>3.00–4.50</td>
</tr>
<tr>
<td>Progestosterone IUD (based on small numbers, thus probably the same as non-contraceptive users)</td>
<td>6.80</td>
</tr>
<tr>
<td>Levonorgestrel IUD</td>
<td>0.20</td>
</tr>
<tr>
<td>TCu-380A IUD</td>
<td>0.20</td>
</tr>
</tbody>
</table>

### Side Effects

With effective patient screening and good insertion technique, the copper and medicated IUDs are not associated with an increased risk of infertility after their removal. Even if IUDs are removed for problems, subsequent fertility rates are normal.

The symptoms most often responsible for IUD discontinuation are increased uterine bleeding and increased menstrual pain. Within one 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. Because bleeding and cramping are most severe in the first few months after IUD insertion, treatment with a nonsteroidal anti-inflammatory (NSAID) agent (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. 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.
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. 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 (e.g., serum ferritin) indicative of iron deficiency. Assessment for iron depletion and anemia should be considered, however, in long-term users and in women susceptible to iron deficiency anemia.

Because of a decidualizing, atrophic impact on the endometrium, amenorrhea can develop over time with the progestin-containing IUDs. With the levonorgestrel IUD, 70% of patients are oligomenorrheic and 30% amenorrheic within 2 years. 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. Sufficient progestin reaches the systemic circulation from the levonorgestrel-containing IUD so that androgenic side effects, such as acne and hirsutism, can occur. More extensive clinical studies are needed to assess the impact of this IUD on the lipoprotein profile; however, it is unlikely that the low dose of levonorgestrel has an important effect on cardiovascular risk.

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 IUD 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. Duration of use does not affect disease 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. 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 prior to MRI, and neither patients nor workers need be excluded from MRIs or the MRI environment.

**Infections**

IUD-related bacterial infection is now believed to be 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. Infections that occur 3–4 months after insertion are believed to be due to acquired STDs, not the direct result of the IUD. The early, insertion-related infections, therefore, are polymicrobial, 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. In nearly 23,000 insertions, however, only 81 cases of PID were diagnosed, and a scarcity of PID was observed in those situations where STDs are rare. There was no statistically significant difference comparing the copper IUD to the inert Lippes Loop or progestin-containing IUDs. These data confirm earlier studies that the risk of infection is highest immediately after insertion and that PID risk does not increase with long-term use. 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.

Doxycycline (200 mg) or azithromycin (500 mg) administered orally one hour prior to insertion can provide protection against insertion-associated pelvic infection, but prophylactic antibiotics are probably of little benefit for women at low risk for STDs.

Compared with oral contraception, barrier methods, and hormonal IUDs, there is no reason to think that nonmedicated or copper IUDs can confer protection against STDs. However, the levonorgestrel-releasing IUD has been reported to be associated with a protective effect against pelvic infection, and the copper IUD is associated with lower titers of anti-chlamydial antibody. In vitro, copper inhibits chlamydial growth in endometrial cells. Thus, the association between IUD use and pelvic infection (and infertility) is now seriously questioned. 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 STD 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 STDs. Women at low risk are unlikely to have pelvic infections while using IUDs.

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. Three studies have attempted to document bacteremia during IUD insertion or removal. Only one of the three could find blood culture evidence of bacteremia, and it was present transiently in only a few patients. In our view, the IUD is acceptable for patients at risk of bacterial endocarditis, but antibiotic prophylaxis (amoxicillin 2 g) should be provided one hour 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 bid 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.

For simple endometritis, in which uterine tenderness is the only physical finding, doxycycline (100 mg bid 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-STD organisms.

**Appropriate outpatient management of less severe infections:**

- Cefoxitin (2 g IM) plus probenecid (1 g orally), or Ceftriaxone (250 mg IM) plus doxycycline (100 mg bid orally), for 14 days.

- Severe infections require hospitalization and treatment with:
  - Cefoxitin (2 g IV q 6 h), or Cefotetan (2 g IV q 12 h)
  - Plus doxycycline (100 mg bid orally or IV)
  - Followed by 14 days of an oral regimen of antibiotics.

The following is an alternative regimen:

- Clindamycin (900 mg IV q 8 h), plus Gentamicin (2 mg/kg IV or IM followed by 1.5 mg/kg q 8 h).

There is a suggestion (in cross-sectional studies) that IUD use increases the risk of HIV transmission from man to woman, especially when the IUD is inserted or removed during exposure to the infected man. 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. In the only study reported, no evidence for female-to-male HIV transmission with IUD use was detected. HIV-infected women who utilize IUDs for contraception do not have a greater incidence of complications (including pelvic
Inflammatory disease. 10

Actinomyces

The significance of actinomycesis infection in IUD users is unclear. There are several reports of IUD users with unilateral pelvic abscesses containing Actinomyces. 11 12 13 However, Actinomyces are found in Pap smears of up to 30% of plastic IUD wearers when cytologists take special care to look for the organism. The rate is much lower (less than 1%) with copper devices and varies with duration of use. 14 15 16 17 Furthermore, actinomyces are commonly present in the normal vagina. 18 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 mgqid that should continue for a month. 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. Another anaerobic, gram-positive rod, Eubacterium nodatum, resembles Actinomyces and has also been reported to be associated with colonization of an IUD. 19 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. 20 After removal of an IUD with visible strings, the spontaneous miscarriage rate is approximately 30%. 21 22 Combining ultrasonography guidance with carbon dioxide hysteroscopy, an IUD with a missing tail can be identified and removed during early pregnancy. 23 If the IUD is easily removed without trauma or expelled during the first trimester, the risk of spontaneous miscarriage is not increased. 24 25 26

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. 27 28 29 There have been no deaths in the United States since 1977 among women pregnant with an IUD. 14 30

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. 31 32 A case-control study did not find an increased incidence of IUD use in pregnancies resulting in limb reduction deformities. 33 34

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. 35 36 37 38 39

Other Complications

Obstetrical complications at delivery (such as 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 STDs. Age and parity are not the critical factors in selection; the risk factors for STDs 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 disease after IUD insertion. 40 41 42 43 It would be prudent for these women to use condoms for STD 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. 44 45 46 47 Most women are good candidates for the IUD.

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 might benefit from a 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 leiomyoma. According to conventional wisdom, the few infections that have been documented (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 is less than that consumed in a normal diet. 48

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. 49 50 51 52 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. 53 54 55 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. 56 57 58 59 Given appropriate screening, counseling, and care, the IUD can provide long-term effective contraception for adolescents.

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
Protection against unwanted pregnancy begins immediately after insertion.

A woman with a previous ectopic pregnancy can use a copper IUD or the levonorgestrel IUD. Women with heavy menses and dysmenorrhea, including women who have a bleeding disorder or are anticoagulated, should consider a progestin-releasing IUD. Ectopic pregnancies can still occur.

Menses can be longer and heavier (except with hormonal IUDs); tampons can be used. 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 old, 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. 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. Postvacinal delivery insertion is not reported 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.

Insertion of an IUD in breastfeeding women is relatively easy, and is associated with a lower removal rate for bleeding or pain. Reports disagree whether perforation is more common in lactating women.

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.

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 day of the cycle.

Technique for the TCu-380A and the Progestasert

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 chlorhexidine or povidone-iodine. Leave the anti-septic-soaked cotton-tipped applicator in the cervical canal during the procedures prior to insertion of the IUD. Place a paracervical block by injecting one 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 minutes 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 anesthesia and with administration of a nonsteroidal anti-inflammatory agent 30 minutes to one hour 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.

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 either by a sliding plastic flange (TCu-380A) or printed graduations (Progestasert). 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 central regions of the endometrial cavity. Placement in the vertical plane increases the risk of expulsion and pregnancy. To release the Progestasert, remove the thread-retaining plug, and withdraw the insertor tube. 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 insertor 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 insertor 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. With the Progestasert, the shorter string verifies correct placement; the longer thread should be trimmed.

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. Patients should return within 3 months, preferably after the first menses, to confirm the presence of the IUD and to provide support, because bleeding changes and expulsion are most likely to occur during this time. Caution the patient that the first 2 menses are typically heavier. As with all office procedures, patients should be provided a 24-hour phone number for urgent questions or concerns, and especially to report unusual pain, bleeding, or vaginal discharge.

Prophylactic Antibiotics

Doxycycline (200 mg) administered orally one hour prior to insertion can provide protection against insertion-associated pelvic infection, but three double-blind randomized studies, two conducted in Africa and one in Turkey, found no significant advantage in the treated groups. Azithromycin in a dose of 500 mg has also been used prophylactically, presumably offering more protection because of a longer half-life. 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 hour before IUD insertion. In women at low risk for STDs, the incidence of infection is so low that there is little benefit to be expected with prophylactic antibiotics.

Summary: IUD Use and Medical Conditions

1. A woman with a previous ectopic pregnancy can use a copper IUD or the levonorgestrel IUD.
2. Women with heavy menses and dysmenorrhea, including women who have a bleeding disorder or are anticoagulated, should consider a progestin-releasing IUD.
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.

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 para-cervical 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 assing 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. Pregnancy outcome after IUD removal is associated with a normal incidence of spontaneous miscarriage and ectopic pregnancy.

If a patient wishes to continue use of an IUD, a new device can be immediately removed 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. 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. Open the forceps widely and see if the IUD moves when the forceps close on it. If it moves, close the forceps tightly and extract the IUD. If unsuccessful, re-introduce the forceps in a different plane, keeping one jaw of the open forceps firmly against 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, besides 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. Although inert perforated devices without closed loops were previously allowed to remain in the abdominal cavity, current practice is to remove any perforated IUD. Because IUD perforations usually occur at the time of insertion, it is important to check for correct position before 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 medicated), adhesion formation appears to be an immediate reaction which does not progress, and rarely leads to serious complications. In appropriate situations (where the risk of surgery is considerable), clinician and patient may elect not to remove the translocated IUD. 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 upon diagnosis.

IUD Myths

We hope the information in this chapter will lay to rest 4 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. The modern IUD HAS NOT exposed clinicians to litigation.
4. IUDs DO NOT increase the risk of ectopic pregnancy.

The IUD for Older Women

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 STDs is 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.

Chapter References

The use of vaginal contraceptives is probably as ancient as homo sapiens. References to sponges and plugs appear in the earliest of writings. Substances with either barrier or spermicidal properties (or both) have included honey, alum, spices, oils, tannic acids, lemon juice, and even crocodile dung. 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 gymnike 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 physicians 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 BC 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. 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 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. Gabriello Fallopius, one of the early authorities on syphilis, described, in 1564, 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.

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.

Vulcanization of rubber dates to 1844, and by 1850, rubber condoms were available in the U.S. The vulcanization of rubber revolutionized transportation and contraception. The introduction of liquid latex and automatic machinery ultimately made reliable condoms both plentiful and affordable.

Diaphragms first appeared in publication in Germany in the 1880s. A practicing German gynecologist, C. Haase, wrote extensively about his diaphragm, using a pseudonym of 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 towards 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. 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 douche syringes were widely advertised; however, they were not widely utilized. 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. Called "Dr. Reid's Clinic,” women of Boston were able to receive contraceptives not available elsewhere in the city.

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 pre-arranged scenario scripted by Buxton and Griswold at the invitation of the district attorney. Found guilty, and fined $100, 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.
The diaphragm made with a flat metal spring, 5 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 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

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 Possibly, the rim of the diaphragm presses against the urethra and causes irritation which is perceived as infectious in origin, or true infection may result from touching the perineal area or incomplete emptying of the bladder. It is more probable that spermicides used with the diaphragm can increase the risk of bacteriuria with E colii, perhaps due to an alteration in the normal vaginal flora. 25 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 trimethoprim-sulfamethoxazole (1 tablet postcoitus), nitrofurantoin (50 or 100 mg postcoitus), or cephalexin (250 mg postcoitus).

Improper fitting or prolonged retention (beyond 24 hours) can cause vaginal abrasion or mucosal irritation. There is no link between the normal use of diaphragms and intrauterine devices, and fell to about 1.9% in 1995. (Chapter 21).

Risks and Benefits Common to all Barrier Methods

Barrier (condoms and diaphragms) and spermicide methods provide protection (about a 50% reduction) against sexually transmitted diseases (STDs) and pelvic inflammatory disease (PID). 3 5 14 17 This includes infections with chlamydia, gonorrhea, herpes simplex, cytomegalovirus, human papillomavirus, and human immunodeficiency virus (HIV); however, only the condom has been proven to prevent HIV infection. STD protection has a beneficial impact on the risk of tubal infertility and ectopic pregnancy. 6 7 There have been no significant clinical studies on STDs 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. 14 27 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. 27 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. 27 This was not confirmed in a careful analysis of two large prospective pregnancy studies. 27 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 v. 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. (Chapter 21).

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 one year of use is 18–20%. 14 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. 27

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 Possibly, the rim of the diaphragm presses against the urethra and causes irritation which is perceived as infectious in origin, or true infection may result from touching the perineal area or incomplete emptying of the bladder. It is more probable that spermicides used with the diaphragm can increase the risk of bacteriuria with E colii, perhaps due to an alteration in the normal vaginal flora. 25 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 trimethoprim-sulfamethoxazole (1 tablet postcoitus), nitrofurantoin (50 or 100 mg postcoitus), or cephalexin (250 mg postcoitus).

Improper fitting or prolonged retention (beyond 24 hours) can cause vaginal abrasion or mucosal irritation. There is no link between the normal use of diaphragms and the toxic shock syndrome. 27 It makes sense, however, to minimize the risk of toxic shock by removing the diaphragm after 24 hours and during menses.

Benefits

Diaphragm use reduces the incidence of cervical gonorrhea, 2 pelvic inflammatory disease, 21 and tubal infertility. 21 This protection may be due in part to the simultaneous use of a spermicide. There are no data, as of yet, regarding the effect of diaphragm use on the transmission of the AIDS virus (HIV). 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 types of 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 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

---

### Table: Combined Contraception Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Combination Pill</th>
<th>Progestin only</th>
<th>IUDs</th>
<th>Levonorgestrel IUD</th>
<th>Copper T 380A</th>
<th>Norplant</th>
<th>Female sterilization</th>
<th>Male sterilization</th>
<th>Depo-Provera</th>
<th>Spermicides</th>
<th>Periodic abstinence</th>
<th>Ovulation method</th>
<th>Symptothermal</th>
<th>Post-ovulation</th>
<th>Withdrawal</th>
<th>Cervical cap</th>
<th>Nulliparous women</th>
<th>Parous women</th>
<th>Nulliparous women</th>
<th>Parous women</th>
<th>Diaphragm and spermicides</th>
<th>Condom</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.5</td>
<td>1.5</td>
<td>0.6</td>
<td>0.1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.1</td>
<td>0.3</td>
<td>6.0</td>
<td>9.0</td>
<td>20.0</td>
<td>3.0</td>
<td>1.0</td>
<td>4.0</td>
<td>26.0</td>
<td>20.0</td>
<td>28.0</td>
<td>18.0</td>
<td>6.0</td>
<td>20.0</td>
<td>20.0</td>
<td>14.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

---

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 v. 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. (Chapter 21).

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 one year of use is 18–20%. 14 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. 27

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 Possibly, the rim of the diaphragm presses against the urethra and causes irritation which is perceived as infectious in origin, or true infection may result from touching the perineal area or incomplete emptying of the bladder. It is more probable that spermicides used with the diaphragm can increase the risk of bacteriuria with E colii, perhaps due to an alteration in the normal vaginal flora. 25 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 trimethoprim-sulfamethoxazole (1 tablet postcoitus), nitrofurantoin (50 or 100 mg postcoitus), or cephalexin (250 mg postcoitus).

Improper fitting or prolonged retention (beyond 24 hours) can cause vaginal abrasion or mucosal irritation. There is no link between the normal use of diaphragms and the toxic shock syndrome. 27 It makes sense, however, to minimize the risk of toxic shock by removing the diaphragm after 24 hours and during menses.

Benefits

Diaphragm use reduces the incidence of cervical gonorrhea, 2 pelvic inflammatory disease, 21 and tubal infertility. 21 This protection may be due in part to the simultaneous use of a spermicide. There are no data, as of yet, regarding the effect of diaphragm use on the transmission of the AIDS virus (HIV). 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 types of 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 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 in order 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, the fit to be 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 hours 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 hours (but no more than 24 hours) 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 re-introduction into the United States. There are several types of cervical caps, but only the cavity rim (Pristic) cap is approved in the U.S. U.S. trials have demonstrated the cervical cap to be 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). Efficacy is significantly reduced in parous women.

The cervical cap has several advantages over the diaphragm. It can be left in place for a longer time (up to 48 hours), and it need not be used with a spermicide. However, a tablespoonful of spermicidal cream or jelly should be placed in the cavity 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 hours).

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 4 sizes 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 it is covered.

To remove the cap (at least 8 hours 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 cervical cap can be left in place for 2 days, but some women experience a foul-smelling discharge by 2 days. Like the diaphragm, it must be left in place for at least 8 hours after sexual intercourse in order to ensure that no motile sperm are left in the vagina.

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 mucus. It seems likely (although not yet documented) that cervical caps would provide the same protection from sexually transmitted diseases as the diaphragm.

The Fem Cap, made of nonallergic silicone rubber, is shaped like a sailor's hat, a design that allows a better fit over the cervix and in the vaginal fornices. This cap may be easier to fit and use, and provides better efficacy. There are two sizes, one for nulliparous women and a larger size for women who have had a vaginal delivery.

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 one gram of nonoxynol-9. Approximately 20% of the nonoxynol-9 is released over the 24 hours the sponge is left in the vagina. Production of the "Today" sponge in the U.S. ceased in 1995, and availability awaits a new manufacturer. "Protectaid" is a polyurethane sponge available in Canada and Hong Kong that contains 3 spermicides and a dispersing gel. The spermicidal agents are sodium cholate, nonoxynol-9, and benzalkonium chloride. This combination exerts antiviral actions in vitro. The dispersing agent, polydimethylsiloxane, 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 hours after sexual intercourse before removal, even if the sponge has been in place for 24 hours before intercourse (maximal wear time, therefore, is 30 hours). It can be inserted immediately before sexual intercourse or up to 24 hours 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 hours before intercourse, and it is easier to remove than the Today sponge. Obviously, the sponge is not a good choice for women with anatomical 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. Some studies indicated higher failure rates (twice as high) in parous women, suggesting that one size may not fit all users.
Discontinuation rates are generally higher among sponge users, compared to diaphragm and spermicide use. For some women, however, the sponge is preferred because it provides continuous protection for 24 hours 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.

**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 sulfate. 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 (this same action occurs with bacteria and viruses, explaining the protection against STDs). 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%.

<table>
<thead>
<tr>
<th>Representative Products:</th>
<th>VCF (70 mg nonoxynol-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaginal Contraceptive Film</td>
<td>Delfen (nonoxynol-9, 12.5%)</td>
</tr>
<tr>
<td>Foams</td>
<td>Emko (nonoxynol-9, 8%)</td>
</tr>
<tr>
<td>Jellies and Creams</td>
<td>Koromex (nonoxynol-9, 12.5%)</td>
</tr>
<tr>
<td>Suppositories</td>
<td>Encare (nonoxynol-9, 2.27%)</td>
</tr>
<tr>
<td></td>
<td>Koromex Inserts (nonoxynol-9, 125 mg)</td>
</tr>
<tr>
<td></td>
<td>Semicid (nonoxynol-9, 100 mg)</td>
</tr>
</tbody>
</table>

“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 hours. Although available without prescription, adequate clinical trial data are not available. Allendale-N9 is a new vaginal contraceptive film that contains more nonoxynol-9 than VCF. An Allendale film has also been developed that contains benzalkonium chloride instead of nonoxynol-9. In addition to spermicidal activity, benzalkonium chloride is microbicidal and demonstrates activity against HIV. Benzalkonium chloride is available for contraceptive use in the form of a suppository, in a sponge, or as a cream in several countries.

**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. There are no comparative studies to indicate which preparations, if any, are better or worse.

Spermicides require application 10–30 minutes prior to sexual intercourse. Jellies, creams, and foams remain effective for as long as 8 hours, but tablets and suppositories are good for less than one hour. 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 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.

Spermicides provide protection against sexually transmitted diseases. In vitro studies have demonstrated that contraceptive spermicides kill or inactivate most STD pathogens, including HIV. Spermicides have been reported to prevent HIV seroconversion as well as to have no effect. Clinical studies indicate reductions in the risk of gonorrhea, pelvic infections, and chlamydial infection. There is little difference in the incidence of trichomoniasis, candidiasis, or bacterial vaginosis among spermicide users. It is probable that spermicides do not provide additional protection against STDs over that associated with condoms, and, therefore, spermicides should not be used without condoms if a primary objective is to prevent infection with HIV, gonorrhea, or chlamydia.

**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 analysis, including a meta-analysis, concluded that there is insufficient evidence to support these associations. Spermicides are not absorbed through the vaginal mucosa in concentrations high enough to have systemic effects. 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.

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 flora promoting the colonization of E. coli, leading to a greater susceptibility to urinary tract infections as noted with diaphragm/spermicide users.

**The Search for Contraceptives to Prevent STDs**

Research is underway to develop microbicides and contraceptives to prevent STDs. The ideal agent would be a topical microbicide that would prevent infection and be spermicidal. 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. However, it is unlikely that any new agent can match the latex condom, which is nearly 100% effective in blocking bacteria and viruses.

**Condoms**

Although awareness of condoms as an effective contraceptive method as well as protectors against STDs has increased tremendously in recent years, a great deal remains to be accomplished in order to reach the appropriate level of condom use. Contraceptive efficacy and STD 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 early 2000s 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" (lambs' 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 which cause STDs and AIDS also do not penetrate latex condoms, but they can penetrate condoms made from intestine. Condom use (latex) also probably prevents transmission of human papillomavirus (HPV), because the conglobulata acuminata. Because spermicides provide significant protection against STDs, condoms and spermicides used together offer more protection than either method used alone. The use of spermicides or spermicide-coated condoms, however, increases the incidence of E. coli bacterium and urinary tract infections in women because of the spermicide-induced alteration in vaginal flora. 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 to 12.7% conversion after 24 months in couples with inconsistent use. Women who are partners of condom users are less likely to be HIV-positive. Polyurethane condoms are expected to protect against STDs and HIV, based upon 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. However, in a randomized, well-designed study, the polyurethane condom had a 6-fold higher breakage rate. 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. Condoms that incorporate a spermicidal agent coating the inner and outer surfaces logically promise greater efficacy, but this remains to be determined. Some women are allergic to the spermicide, and it is a concern (not documented) that mucosal lesions can promote HIV transmission should a condom break.

An often repeated concern is the alleged reduction in penile glans sensitivity that accompanies condom use. 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 obstruction 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 STDs. 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 stiff 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.

Summary — Key Steps for Maximal Condom Efficacy

1. Use condoms for every act of coitus.
2. Place the condom before vaginal contact.
3. Withdraw while the penis is still erect.
4. Hold the base of the condom during withdrawal.
5. Use a spermicide or a condom lubricated with a spermicide.

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, and 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, and 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%. With experienced couples, condom failure due to breakage and slippage (sufficient to increase the risk of pregnancy or STDs) occurs at a rate of about 1%. In a U.S. survey, one pregnancy resulted for every 3 condom breakages. Concomitant use of spermicides lowers failure rates in case of breakage.

Breakage is a greater problem for couples at risk for STDs. An infected man transmits gonorrhea to a susceptible woman approximately two-thirds of the time. If the woman is infected, transmission to the man occurs one-third of the time. The chances of HIV infection after a single sexual exposure range from one in 1000 to one in 10. 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 within 72 hours. 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 supply of oral contraceptives should be kept available for self-medication (see Chapter 22).

For the immediate future, prevention of STDs 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 STD prevention. The main motivation for condom use among women continues to be prevention of pregnancy, not prevention of STDs. An important area of concentration is the teaching of the social skills required to ensure use by a reluctant partner. We believe that bars on condom advertising should be eliminated. Using scare tactics about STDs 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 STDs will preserve future fertility.

We suggest that clinicians consider making free condoms available in their office. Manufacturers will sell condoms at a bulk rate, from $50—$100 per 1000, depending on style and lubrication.

U.S. CONDOM MANUFACTURERS:
Ansell Health Care (telephone: 800–327–8659)
Carter Products (telephone: 609–655–6000)
Meyer Laboratories (telephone: 800–426–8366)
Okamoto USA (telephone: 800–283–7546)
Schmid Laboratories (telephone: 800–829–0867)

The Female Condom
The female condom is a pouch made of polyurethane, which lines the vagina. 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 lubricated with silicone, and a sponge or applicator need not be used. The female condom should be an effective barrier to STD infection; however, high cost and acceptability are major problems. The devices are more cumbersome than condoms, and studies have indicated relatively high rates of problems such as slippage. Women who have successfully used barrier methods and who are strongly motivated to avoid STDs 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.

**Future Developments**

New barrier devices are being pursued, such as sponges incorporating several spermicides and cervical caps made of different materials. Chemical agents are being investigated that can combine spermicidal and antimicrobial actions, and vaginal spermicidal films of different materials and containing other spermicidal agents are being tested. Disposable diaphragms that release spermicide are in development.

Lea's Shield is a vaginal barrier contraceptive composed of silicone. 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 to enhance STD protection. Lea's Shield is designed to remain in place for 48 hours after intercourse. Pregnancy rates are similar to other barrier methods, and no serious adverse effects have been reported.

**Chapter References**

73. Hunt WL, Gabby L, Potts M, Lea's Shield®, a new barrier contraceptive; preliminary clinical evaluations, three-day tolerance study, Contraception 50:551, 1994.
**Female Infertility**

*The Epidemiology of Infertility in the U.S.*

**Concern With Infertility**

- Aging and fertility
- Endocrine Changes With Aging
- Assisted Reproduction in Older Couples
- The Role of the Physician
- The Female Infertility Investigation
- The Proceptive Test
- Uterine Endometriosis
- Tubal Occlusion
- Luteal Failure
- Luteal Phase Defect (Inadequate Luteal Phase)
- Luteinized Unruptured Follicle
- Myomas
- Endometriosis
- Venereal Infertility
- Drugs
- Myths and Aseotropes Coupled

**Chapter References**

Infertility is defined as one year of unprotected coitus without conception. It affects approximately 10–15% of couples in the reproductive age group, which makes it an important component of the practices of many physicians. Fecundability is the probability of achieving a pregnancy within one menstrual cycle (about 25% in normal couples); fecundity is the ability to achieve a live birth within one menstrual cycle.

There have been 3 striking changes in infertility practice during the past 2 decades. First was the introduction of in vitro fertilization and other assisted reproductive technologies (ART) that have enlarged the possibilities for successful treatment and provided an opportunity to study basic reproductive processes. ART refers to all techniques involving direct retrieval of oocytes from the ovary (Chapter 31). Second, and partially because of the media attention focused on ART, the public has become more aware of potential treatments, and this has generated a marked increase in patient visits for infertility. The third change is the increase in the proportion of women over 35 seeking medical attention for infertility. One of every five women in the United States is having a first child after 35, a marked increase over earlier figures. This reflects both a later age for marriage and postponement of pregnancy in marriage as women, by choice or by circumstances, commit to the work place. There also has been a slight increase in the proportion of couples considered infertile. More importantly, there is an increasing number of infertile couples due to the aging of the large post World War II population boom generation.

**The Epidemiology of Infertility in the U.S.**

The first United States census was in 1790. At that time, the birth rate was 55 per 1000 population; 200 years later, it is 15.5 per 1000 population, a decrease from 8 births per woman to 1.2. There are some obvious and some speculative explanations for the decline in U.S. fertility.

**Popular Explanations for the Decline in U.S. Fertility**

- Changing roles and aspirations for women.
- Postponement of marriage.
- Delayed age of childbearing.
- Increasing use of contraception.
- Liberalized abortion.
- Concern over the environment.
- Unfavorable economic conditions.

The deferment of marriage and postponement of pregnancy in marriage are the most significant changes in modern society. Only 16% of the decline in the total fertility rate in the U.S. is accounted for by the increase in the average age at first marriage; 83% of the decline in the total fertility rate is accounted for by a change in marital fertility rates (avoiding pregnancy in the first years of marriage).

After World War II, the U.S. total fertility rate reached a modern high of 3.8 births per woman. The last women born in this period won’t be reaching their 45th birthdays until around 2010. For the next decade, therefore, there will continue to be an unprecedented number of women in the later child-bearing years. The aging of the World War II population boom is giving current times a greater number of women who are delaying marriage and childbirth. The proportion of births accounted for by this older group of women increased by about 72% from 1982 to 2000. These couples, pressed for time, have a desire to get pregnancies accomplished in a shorter time period. In 1996, the U.S. birth rate for women aged 30–34 was the highest since 1966, for women aged 35–39, the highest since 1968, and for women aged 40–44, the highest since 1971.

**Concern With Infertility**

In 1995, approximately one in 6 women of reproductive age reported that they had sought professional help during their lifetimes because of infertility. A sharp escalation of demand for infertility services began in 1981. From approximately 600,000 visits in 1988, the total increased to nearly 1 million in the early 1970s, then in the early 1980s, the total went over 2 million visits. In 1995, 15% of reproductive age women (9.3 million) reported some use of an infertility service, compared with 12% (6.6 million women) in 1988. Approximately 1% of 60.2 million women of reproductive age used one of the methods of assisted reproductive technology, and 3% used ovulation drugs.

There was no significant change in the proportion of U.S. infertile couples in the 1980s. However, the overall percentage of infertile women increased from 8.4% in 1982 and 1988 to 10.2% (6.2 million) in 1995. Although this increase is significantly due to the increase in older women who were born in the years after World War II, there is also an increase across all age groups, partly reflecting the effect of sexually transmitted diseases and partly an increasing concern for infertility.

Why then is there this increasing concern for infertility? First, there is an increasing number of infertile couples. As noted, the aging of the post World War II generation is yielding a greater number of women who are delaying marriage and childbirth. From 1982 to 1995, the number of childless women aged 35–44 years increased by about 1 million, from 1.9 million in 1982 to 2.8 million in 1995. In addition, there is a greater awareness of modern treatments and a greater ability to afford health care. The impact of widespread publicity is significant. There is an increased availability of services, and there is a greater knowledge of the diagnosis and management of infertility among clinicians. Finally, infertility is now more socially acceptable as a problem.

The post World War II generation has faced a unique evolutionary change. They were the first to be able to exercise control over their fertility, and then as they aged and deferred pregnancy, they had to deal with the problem of unintended infertility. Because many couples defer pregnancy and then desire their families within a condensed interval of time, there is a sudden urgency regarding fertility. These factors have combined to substantially increase the number of couples seeking and using infertility services. Nevertheless, in the 1995 U.S. National Survey of Family Growth, the majority of women (56%) with impaired fecundity had not obtained any
Aging and Fertility

Age alone impacts on fertility; aging of the reproductive system plays a role and spontaneous abortion provides another factor. The majority of early spontaneous abortions after age 35 are due to autosomal trisomies, the incidence of which increases with maternal age. The risk of clinically recognized spontaneous abortion increases from about 10% under age 30, to 18% in the late 30s, and 34% in the early 40s. Indeed, once pregnancy is achieved by an older woman, the greatest obstacle to successful delivery is the risk of spontaneous loss. The frequency of both euploid (normal) and aneuploid (abnormal) abortuses increases with maternal age. The overall abortion risk (recognized and unrecognized) in women over age 40 is approximately 75%. Studies of oocytes from older women have suggested that older oocytes have impaired meiotic spindle formation or function that could cause the failure of proper chromosome migration.

In addition, as women enter their 30s, there is a greater likelihood of being affected by a number of diseases, for example endometriosis, that can interfere with fertility. Cumulative exposures to occupational or environmental hazards also could lessen fertility as a woman ages. Another factor that has contributed to infertility at all ages is the spread of sexually transmitted diseases with their damaging effect on the fallopian tubes.

What is the exact impact of age on fertility? The classic study is that of the Hutterites who live in the Dakotas, Montana, and the adjacent parts of Canada. The sect originated in Switzerland in the 1500s, and practically all of the living Hutterites came to South Dakota in the 1870s. Four colonies of 443 Hutterites grew, by 1950, to 93 colonies containing 5,452 Hutterites. Contraception is condemned, and because of the communal arrangement of their society, there is no incentive to limit the size of families. All families are provided for equally.

In the 1950s, Christopher Tietze analyzed the fertility rates of the Hutterites. Only 5 of 209 women had no children for an infertility rate of 2.4%. The average age of the women at the last pregnancy was 40.9 years, and there was a definite decrease in fertility with age. Eleven percent of the women bore no children after age 34; 33% of the women were infertile by age 40, and 87% were infertile at age 45.

The fertility of the Hutterites has become legendary. Their fertility rate is an example of how high fertility can be when a population is healthy, stable, and not using contraception. Based on the Hutterite data, it can be concluded that a population that marries around age 22, has some lactational amenorrhea, and some age-related and parity-related decline in coital frequency, can produce 11 live births per married woman. Hypothetically, if marriage were earlier, there were no lactational amenorrhea, and no sterilization or decline in coitus, the total fertility rate would be about 15 live births per woman.

The French studied the pregnancy rate in a donor insemination program, including only women with azoospermic husbands. These women were less likely to have infertility factors than those women married to oligospermic males. Below the age of 31 the pregnancy rate over one year was 74%. This decreased to 62% at ages 31 to 35 and to 35% when the women were older than 35. Because donor insemination was used, the effect of reduced coital frequency in older couples did not influence this study. A recent prospective study with donor insemination has demonstrated a similar relationship with age. Of note was the requirement for more treatment cycles to achieve pregnancy in older women; 9–10 cycles rather than the usual 6. In a donor insemination program in the Netherlands, the probability of having a healthy baby decreased 3.5% per year after age 30. A woman aged 55% had 50% of a chance of having a healthy baby compared to a woman aged 25. In a study of almost 3000 donor insemination cycles in one center, the cumulative conception rates after 3, 6, and 12 cycles were 21%, 40%, and 82% for patients less than 30 years old, compared with 17%, 26%, and 44% for women 30 years and older.

The decline of fertility among married couples with advancing age has been repeatedly documented. It is safe to say that about one-third of women who defer pregnancy until the mid to late 30s will have an infertility problem, as will at least half of women over age 40. In programs of assisted reproductive technologies, delivery rates in women over age 40 are one-third to one-half of those in younger women, and many programs have only rare successful pregnancies in women over age 40.

The oldest spontaneous pregnancy in modern times (according to The Guinness Book of World Records) occurred in a woman from Portland, Oregon, who delivered when she was 57 years and 120 days old. In older times, a Scottish woman was reputed to have delivered 6 children after the age of 47, the last at age 62!

There are changes in the male reproductive system with aging. A decline in testosterone and an increase in gonadotropins are associated with a decrease in sperm production and the number of normal sperm. There are reasons to believe that the quality of sperm decreases. A paternal age greater than 40 is associated with a 20% greater chance of birth defects in the offspring. These changes in the male with aging are modest, but significant, although the capacity to fertilize is maintained. A short report in 1935 told of a North Carolina farmer born in 1840. Over a period of 30 years, his first wife had 16 children. At age 53, he remarried to a 27-year-old widow; one year later, a child was born. Reading the detailed physical examination provided in this report, it is obvious that this 94-year-old man was very unattractive, with leathery, wrinkled skin and no teeth. It is by no means certain that he was the father of the child.

Endocrine Changes With Aging

The early fetal mitotic multiplication of germ cells produces a total of 6–7 million oogonia by 18–20 weeks of pregnancy. From this point in time, the germ cell content will irretrievably decrease. At the onset of puberty, the germ cell mass has already been reduced to approximately 300,000 units. Over the next 35–40 years of reproductive life, during which time only 400–500 oocytes will ovulate, these units will be depleted by atresia to a point at menopause when only a few hundred remain.

In the last 10–15 years before menopause, there is an acceleration of follicular loss. This accelerated loss begins when the total number of follicles reaches approximately 25,000, a number reached in normal women at age 37–38. This loss correlates with a subtle but real increase in follicle-stimulating hormone (FSH) and decrease in inhibin. These changes reflect the reduced quality and capability of aging follicles. These subtle changes (which influence fertility) are not associated with any noteworthy or observable changes, other than an alteration in menstrual cycle characteristics.

Prior to the menopause, there is a period with shorter follicular phases, with increased FSH levels, but normal luteinizing hormone (LH) levels and luteal phases. The menstrual periods for 10–15 years before the menopause are regular, but there is a steady decrease in cycle length due to the shortened follicular phase. 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. For a period of several years (as much as 10 years in some women) prior to menopause, the cycles lengthen again.

Women who present with "incipient" ovarian failure have elevated FSH levels and decreased levels of inhibin, but normal levels of estradiol. The rise in FSH during the later years is in response to declining inhibin production. Inhibin levels are lower in the follicular phase in women 45–49 years old compared to younger women. The specific inhibin involved in this feedback effect on FSH during the follicular phase is inhibin B. In addition, a decline in luteal phase inhibin A can contribute to a greater rise in FSH during the luteal phase transition. This decline in inhibin begins early, but accelerates after 40 years of age. The rise in FSH occurs as soon as age 20, but there is no change in luteinizing hormone (LH) levels until menopause. 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). Arguing in favor of a role for a reduced follicular pool is the observation that follicular fluid obtained from prevulatory follicles of older women contains amounts of inhibin A and B that are similar to that measured in follicular fluid from young women.

These endocrine changes and the decline in ovarian follicles are correlated with a decrease in ovarian volume. A measurement of volume by ultrasonography that is equal to or less than 3 cm³ predicts a poor response to ovulation induction.

Testing the Ovarian Reserve

The measurement of baseline hormonal levels provides important and helpful prognostic information. Because FSH levels increase with age, elevated FSH levels on cycle day 3 (greater than 12 IU/L, but especially greater than 20 IU/L) are associated with poor performance with in vitro fertilization. A cycle day 3 FSH level that is 25 IU/L or more, or an age of 44 years or more, are both independently associated with a chance of pregnancy close to zero during ovulation induction or with assisted reproductive technologies. Values can vary between laboratories, significantly due to differences in antibodies and standards used to measure.
There is a high incidence of FSH levels. With the most recent assays, day 3 values above 10 IU/L are abnormal and are comparable with values greater than 20 IU/L with older assays.

Women with one ovary have higher day 3 FSH levels that correlate with reduced outcomes in in vitro fertilization. It is not certain whether this reflects the loss of the other ovary or the factors that were responsible for the unilateral oophorectomy. Keep in mind that there is no abrupt change at 40, and therefore, these changes can apply to younger women. The age of change is determined by the rate of oocyte loss, a rate that is genetically preprogrammed in most women. An elevated FSH at any age is correlated with fewer available oocytes, and advancing age adds the factor of lesser oocyte quality. An elevated estradiol level on day 3 (greater than 80 pg/mL) is also predictive of difficulty in achieving pregnancy. Premature estradiol elevations are associated with a hurried recruitment of follicles in response to increased FSH secretion, changes characteristic of the years preceding menopause. When both FSH and estradiol are elevated on day 3, ovarian response to stimulation is very poor.

The Clomiphene Challenge Test. Navot and colleagues developed the clomiphene challenge test as a bioassay of FSH response (which in turn reflects ovarian follicular inhibit capability). Clomiphene is administered in a dose of 100 mg/day on days 5–9. The level of FSH on day 10 is compared with the baseline level on day 3. An exaggerated FSH response to 26 IU/L or more was two standard deviations above the control values in this study, and this increase in FSH was associated with a significant prospect for failure to achieve pregnancy. Remember that there is a wide variety of FSH assays (the newer assays provide numbers that are approximately 70% of the older results), and clinicians must become familiar with the values provided by their own laboratories. There is a high incidence of abnormal responses in women over age 35, and 85% of women with increased FSH levels respond poorly to ovarian stimulation. In a general infertility population, about 10% of women have an abnormal response to the clomiphene challenge test. As expected, the percentage increased with increasing age (from 3% under age 30 to 26% over age 39), and women with unexplained infertility had a 38% incidence of abnormal tests. This test of ovarian reserve is more sensitive than just measuring basal FSH. Ovarian reserve is defined as the ability of the ovary to produce an adequate number of oocytes. When the clomiphene challenge test indicates a poor prognosis no matter the woman’s age. However, because age alone is an important factor (presumably due to other age-related oocyte conditions of quality), older women with normal hormone levels and normal responses to clomiphene still have diminished chances of pregnancy.

We believe it is worthwhile to screen all infertile women beginning at age 30, women of any age with unexplained infertility, and women who respond poorly to ovulation induction with the clomiphene challenge test and a day 3 estradiol level. For practical purposes, equivalent results can be expected on days 2, 3, and 4 of the cycle. Far less certain is the need for these tests in the woman over 40, even though normal tests do not preclude achievement of an ovarian reserve. An abnormal test in a woman over 40, however, can be used to bolster advice to consider ovum donation. Although an abnormal test at any age indicates a very poor prognosis, it does not indicate an absolute inability to achieve pregnancy. Testing the ovarian reserve is just another part of the decision making process for patients and clinicians.

Assisted Reproduction in Older Couples

Experience has repeatedly demonstrated a reduction in I/VF pregnancy rates when the oocytes are of advanced age. When embryos from the same cohort of young donated oocytes were simultaneously transferred to young and older recipients, however, the pregnancy rates were similar. The high rate of implantation and pregnancy among older women receiving donated younger oocytes argues that uterine factors are not involved with the decline in fecundity with aging. Uterine factors that could suffer with age include responses of the endometrium to implantation, capillary permeability and hyperemia, and decidual differentiation.

The reproductive endocrine group at the University of Southern California has concluded that the problems of decreased fecundability and increased pregnancy wastage can be attributed to the effect of aging on oocytes. In 100 consecutive patients in an oocyte donor program, there was no decline in success over the age range from 40 to 50. Furthermore, patients have been achieved in women aged 50–62. There are two important observations regarding these older women. First, the older recipients of young oocytes received extensive medical screening, infectious disease screening, reproductive function assessment, and psychological assessment. The good obstetrical outcomes in these patients must reflect, to an important degree, the youthful health of this older group of patients. Second, the stimulation protocol utilized a 100 mg dose of progesterone.

Meldrum reported a lesser percentage of delivered pregnancies in women over 40 compared with women under 40 going through an identical donor oocyte–IVF program. However, he then achieved pregnancy rates in women over 40 similar to those in women under 40 by increasing the progesterone dose in the stimulation protocol from 50 mg to 100 mg per day. Meldrum argued that this age-related decline in endometrial receptivity for progesterone might be due to an age-related decline in estrogen receptors as demonstrated in animal studies. Others have also concluded that hormone treatment produces a better endometrium and corrects any uterine factor. On the other hand, the uterine factor must not be a major obstacle in older women because in a carefully designed study, good quality ova derived from the same cohort yielded similar pregnancy rates in younger and older recipient women (even though the progesterone dose was 50 mg per day). A pregnancy rate in older women of approximately 40–50% per cycle can be achieved in a donor oocyte program. A cumulative birth rate per patient can reach 90% with 4 or more cycles. The experience with donor oocyte programs argues strongly that the age-related decline in fecundity is primarily due to aging oocytes. These results further point out that the increased risks of spontaneous abortion and chromosomal anomalies associated with older age are also due primarily to aging oocytes. In a large series, the rate of spontaneous abortion in recipients correlated with the age of the donors. The abortion rate increased from 14% in recipients who received oocytes from donors aged 20–24 years to 44.5% when the recipient received oocytes from donors older than age 35. Pregnancy wastage directly correlates with the age of the woman who produces the oocytes.

We know that we can establish pregnancy in women over age 50, now we have to consider whether this feat is a wise accomplishment. We have little control over natural events, but moral and social questions do arise when we have the opportunity to assist a woman to achieve a pregnancy. There can be no objection to the use of these procedures to achieve pregnancies in women who undergo ovarian failure at an early age. But when are couples too old to have a child? It can be argued that the risk involved in the pregnancy (a relatively high incidence of hypertension and gestational diabetes) and the risks involved in raising a teenager when you are yourself a senior citizen are issues that patients and clinicians must resolve. On the other hand, couples over 40 may be ideal parents: high degree of motivation, increased maturity and wisdom, better financial circumstances. They are more able and willing to devote attention to their children. Certainly it is a wise precaution to discourage the use of these techniques by women who are not biologically young and healthy no matter what their age. Besides the health and medical circumstances of the individuals involved, it seems appropriate to also consider the provisions for child rearing. In our view, Paulson and Sauer put it well when they stated that “It seems unlikely that these children will be so disturbed that they would prefer not to have been born than to have suffered this indignity.”

Clinicians can appropriately advise older women that there is no time to waste, and serious consideration should be given to an early resort to hormonal stimulation for both oocyte response and support of the endometrium. Older couples should be provided the option of oocyte donation from young donors instead of standard assisted reproductive technologies.

The Role of the Physician

When is a medical success really a success? Approximately one-half of couples presenting after one year of infertility can be expected to become pregnant spontaneously in the following year. In an English study, only 20% of women who had failed to have a birth within the first two years of marriage never had a child. In a lifetime-analysis of 58 unrelated apparently normal infertile couples, 74% were pregnant by two years; however, normal couples achieve this rate in 9 months. Overall, approximately 40% of couples become pregnant after discontinuation of treatment, and 35% of couples never treated can expect to become pregnant. Therefore, achieving pregnancy is not the only measure of success. One of the important missions for the infertility physician is to speed up the period of time required to achieve pregnancy. For couples in their 30s, the recommendation to seek help promptly is valid — the sooner a problem is detected, the better. Because the incidence of spontaneous pregnancy is significant until 3 years have passed, it is appropriate to require 3 years of infertility in women less than 30 years old before making a diagnosis of infertility. Because age is an important determinant, evaluation and therapy should not be deferred in older women.

In response to the needs of infertile individuals, physicians should have four goals in mind:

1. The first goal is to seek out and to correct the causes of infertility. With proper evaluation and therapy, the majority of women will become pregnant.
2. The second goal is to provide accurate information and to dispel the misinformation commonly gained from friends and mass media.

3. The third goal is to provide emotional support for the couple during a trying period. The inability to conceive generates a feeling in many couples that they have lost control over a very significant segment of their lives. The burden is aggravated by the manipulations that couples have to undergo during the infertility investigation, including the need to have intercourse on schedule. Couples need to have an opportunity to ventilate their concerns and dispel some of their fears. A valuable adjunct to the efforts of the physician are support groups for infertile couples such as those organized by RESOLVE. Meetings in groups allow individuals to realize that their problems are not unique, and it enables them to obtain information on how others cope with infertility. It must be emphasized that, while severe anxieties can interfere with ovulation and frequency of intercourse, there is no substantial evidence that infertility is caused by the usual anxieties besetting a couple trying to conceive.

4. An often neglected goal is that of counseling a couple concerning the proper time to discontinue investigation and treatment, to consider adoption, and/or to move on to the assisted reproductive technologies. This is especially important in the 10–15% of couples with no known cause for their infertility.

Counseling also should be an ongoing process during the infertility investigation and treatment. At least every 3 months the physician and couple should review the status of their care and outline the anticipated management; thereby, changes in plans due to medical, emotional, or financial reasons can be instituted in a timely fashion.

The Female Infertility Investigation

Infertility problems, once thought to be refractory to treatment, have progressively responded to newer therapies. Occlusive tubal disease now is bypassed by in vitro fertilization. Even the most unpromising cases of male infertility can be overcome by the intracytoplasmic injection of sperm. The deleterious effects of aging on female reproduction can be neutralized by the use of ovum donation from younger women.

There are, however, many cases of infertility that yield to less intrusive and less expensive therapy. Even this area has undergone major changes, with a decreased emphasis on diagnosis, and rapid movement toward empiric therapy with superovulation and intrauterine insemination (IUI).

We find it very helpful to mail a detailed questionnaire to our patients prior to their initial visit. The questionnaire is very complete, providing information that ranges from previous medical events and sexuality to recreational, social, and vocational activities. Patients often write comments regarding their past history and their feelings that are difficult to express during an office interview.

There are advantages to having the male present during the initial interview. He may contribute valuable historical and attitudinal information. It also gives the physician the opportunity to emphasize that both partners are involved in the infertility investigation. A male who has been acquainted at its inception with the clinician's treatment of the infertility problem will be less reluctant, as time progresses, to ask for clarification of any aspect of the testing and treatment. This can prevent misunderstandings engendered when the male partner's only source of information is the woman. Early in the clinician-couple interaction, frequency of coitus, and possible sexual problems should be ascertained.

Failure to ovulate is the major problem in approximately 40% of female infertility; another 40% is due to tubal pathology and pelvic pathology, and less than 10% is due to problems such as anatomic abnormalities or thyroid disease. It should be noted that induced abortions do not influence subsequent pregnancy rates. Fetal wastage is definitely higher in diethylstilbestrol-exposed women, and while there is still some uncertainty, evidence suggests that primary infertility is also more common. Besides the well-known impact of smoking on pregnancy, there is good evidence that fecundity is reduced in men and women who smoke, and the risk is greater with smoking at an early age. There is a dose-response relationship between the number of cigarettes smoked and how long it takes to achieve pregnancy. Marijuana inhibits the secretion of GnRH and can suppress reproductive function in both men and women, and cocaine use is known to reduce spermogenesis. Studies have failed to confirm an adverse impact of caffeine on fertility, although high levels of intake may be associated with a delay in conception.

Couples need to be aware that there is a normal time requirement to achieve pregnancy. In each ovulatory cycle, normal couples have only about a 30% chance of becoming pregnant. Guttman's classic table has been a standard since 1956.

<table>
<thead>
<tr>
<th>Time of Exposure</th>
<th>% Pregnant</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 months</td>
<td>57%</td>
</tr>
<tr>
<td>6 months</td>
<td>72%</td>
</tr>
<tr>
<td>1 year</td>
<td>85%</td>
</tr>
<tr>
<td>2 years</td>
<td>93%</td>
</tr>
</tbody>
</table>

Because the male factor accounts for approximately 35% of infertility, examination of the semen should be an early diagnostic step in the investigation. If abnormal, further diagnostic procedures in the woman should be deferred until decisions are reached regarding the man (see Chapter 29). If the semen is normal, attention is directed to the woman.

Laboratory testing should be directed by clinical judgment. However, there are a few specific recommendations. Women with a negative rubella titer should be immunized and attempts for pregnancy delayed for 3 months. When personnel are exposed to sperm and/or oocytes, many states now mandate tests for syphilis, antihumanitis B core, hepatitis B surface antigen, antihumanitis C, and antibody to HIV 1/2 (AIDS) and HTLV-I. Screening for cystic fibrosis should be considered.

The Postcoital Test

The postcoital test provides information regarding both the receptivity of cervical mucus and the ability of sperm to reach and survive in the mucus. Estrogen levels peak just prior to ovulation, and this provides maximal stimulation of the cervical glands. An outpouring of clear, watery mucus is fostered, which may be of sufficient quantity to be noted by the woman. Earlier in the cycle, when estrogen output is lower, and starting 2 to 3 days after ovulation when progesterone levels increase and counteract the estrogen, the mucus is thick, viscid, and opaque.

The postcoital test is performed around the time of the expected luteinizing hormone (LH) surge as determined by a previous basal body temperature chart or by the length of prior cycles. Timing also can be obtained with ultrasonography and LH monitoring, but this usually is not necessary. Between 2 and 8 hours after coitus, cervical mucus is removed with a nasal polyp forceps or tuberculin syringe and examined for macroscopic and microscopic characteristics. Attempts to refine the postcoital test by studying individual fractions from different levels in the cervical canal, with emphasis on the sample from the internal os, have not produced convincing evidence of value. Sperm distribution is uniform throughout the cervical canal, and selective sampling at the level of the internal os is not necessary.
less than 2-hour interval between coitus and examination has been recommended as giving maximal information, but this early evaluation may be deceptive because complement dependent reactions in mucus, which can immobilize sperm may not be apparent for a few hours. Others have suggested that a 16- to 24-hour interval provides a better assessment of sperm longevity, and a study has indicated that there is no drop in the number of sperm at any time during the first 24 hours. There are other indications, however, that the number of sperm does decrease after 8 hours, and this is more in keeping with our experience. Therefore, we suggest that the couple have coitus in the morning or late at night, and that the test be performed 2 to 8 hours later. It is also suggested that the couple abstain from intercourse for 48 hours prior to the postcoital test.

The need for scheduling the postcoital test at precise times in the cycle may produce problems for the couple who cannot have sex on demand. This may further burden a couple already troubled by the need to cope with their infertility and the loss of control involved in the infertility investigation. A physician must be sympathetic to this problem, and on occasion, precise timing must be sacrificed and the woman told to come to the office following unscheduled intercourse.

The stretchability (spinnbarkeit) of the mucus at midcycle should be 8–10 cm or more. This characteristic can be assessed as the mucus is pulled from the cervix, or alternatively, by placing the mucus on a slide, covering it with a coverslip and then lifting the coverslip. At midcycle the mucus contains 95–98% water and should be watery, thin, clear, acellular, and abundant. When dried on a slide, it should form a distinct fern pattern.

If the mucus is thick rather than thin, opaque instead of clear, the proximity of the test to ovulation should be determined by the onset of the next period (or by the temperature chart if one is being taken during that cycle). The best time for the postcoital test is the day prior to the first temperature elevation. If poor mucus quality is related to inaccurate timing, the test should be repeated in a subsequent cycle with scheduling based on urinary LH testing. Poor mucus at midcycle can be a physical barrier that decreases sperm penetration. In one study 54% of women with good mucus became pregnant, compared to 37% with poor mucus, a statistically significant difference. In all likelihood some of these poor tests were reflections of inaccurate timing, and pregnancies do occur even with poor mucus at ovulation time. It remains our impression, however, that poor mucus is associated with a decreased chance for fertility.

Treatment of poor mucus has been attempted by giving 0.625 mg of conjugated estrogens daily for the 8 or 9 days preceding the expected time of ovulation. In a 28-day cycle that would be between days 5 and 13. There is no advantage to continuing the hormone treatment through the luteal phase of the cycle. If the initial treatment with estrogen fails to produce a change in the mucus, the dose is increased to 1.25 mg/day. In refractory cases, 5 mg of conjugated estrogens can be given starting a few days prior to expected ovulation. However, evidence for the value of estrogen therapy is lacking, and it has no demonstrable benefit for the poor mucus which is, on occasion, associated with the use of clomiphene citrate. If there is evidence of chronic cervicitis with thick yellowish mucus, culture for chlamydia is important, and, where appropriate, systemic antibiotics should be used. On rare occasions, the cervix is treated with electrocautery or cryosurgery.

The most logical approach to overcome the barrier of thick cervical mucus is the method we prefer, the use of intrauterine insemination (IUI) of washed sperm (discussed below and in Chapter 29). If poor mucus is the sole problem, IUI for 3–4 cycles is associated with a cumulative pregnancy rate of 40–50%.

In addition to providing an evaluation of the mucus, the postcoital test also gives information concerning the male. Absence of sperm requires a review of the couple's coital technique. Repeated cancellations of appointments for the postcoital test may be a clue that there are sexual problems that have not been uncovered by the interview. However, most often the stress of performance for the scheduled postcoital test is responsible for the failure to achieve coitus. Most importantly, absence of sperm necessitates a detailed review of the semen specimen.

One of the most difficult problems in infertility is the postcoital test that repeatedly shows no sperm or only dead sperm despite good mucus. The patient should be cautioned that lubricants, such as K-Y Jelly and Surgilube, have a spermicidal effect in vitro and should not be used by infertile couples. If lubrication is necessary, vegetable oil can be used without interfering with sperm movement. As noted earlier, re-examination of the semen to check sperm count and motility is a necessity if the postcoital test is poor.

If the semen is normal, the pH of the cervical mucus at midcycle should be determined. Good results have been reported using a precoital douche of 1 tablespoon of sodium bicarbonate in 1 quart of water when a poor postcoital test was associated with a pH below 7. Cervical cultures should be obtained for chlamydia if the mucus is yellowish, and sperm antibody testing should be performed in cases where there are no sperm or mostly nonmotile sperm. In addition, sperm antibody testing (Chapter 29) is mandatory when, in a postcoital test with good mucus, the sperm are found shaking in place but not moving progressively. This shaking movement is a common finding in immunologic infertility.

In vitro cross testing, utilizing donor or bovine mucus and donor sperm, can help to determine whether the poor postcoital test is due to factors in the mucus or to defects in the sperm. However, in clinical practice this type of testing has little value because the next step is treatment with IUI regardless of results with cross testing.

A postcoital test cannot be considered a substitute for a semen analysis. While 21 or more sperm/high power field (HPF) is almost always associated with a sperm count above 20 million/mL, the postcoital test gives little information concerning the morphology of sperm in the ejaculate. There are considerably fewer abnormal forms in the cervical mucus compared to the ejaculate. This may represent a filtering effect of the cervical mucus or may indicate that abnormal forms do not have the motility to penetrate the cervical mucus.

The postcoital test has had a role in the infertility investigation for over 100 years. The postcoital examination first received prominence in the late 1800s in the writings of J. Marion Sims who was primarily interested in how long sperm survived in the mucus. Sims' son, Harry, reported to the American Gynecological Society in 1888 the details and purpose of the postcoital test, essentially as still performed today. The major impetus to general clinical use can be attributed to a book published in 1913 by Max Hühner, entitled Sterility in the Male and Female. Hühner, born in Berlin and a graduate of the College of Physicians and Surgeons of Columbia University, was a urologist in New York City. His second book, Sterility in the Male and Female, was published in 1916 and went through 3 editions and a Spanish translation; this was followed by a third book in 1937. Thus, the postcoital test is also known as the Sims-Hühner test.

The postcoital test has always had its detractors because of limited correlations with fertility, the inability to establish universally agreed upon normal values, and controversies concerning the proper treatment of a poor postcoital test. One study could find no difference in the subsequent pregnancy rates among groups having no sperm, no motile sperm, 1–5 motile sperm, 6–10 motile sperm, and 11 or more motile sperm. Another study indicated that there was a statistically significant increase in the percentage of pregnancies only when there were more than 20 sperm/HPF. Moreover, in a study of postcoital tests in fertile couples, 20% had either no sperm or less than one sperm/HPF. A newer argument raised against use of the postcoital test is that widespread use of intrauterine inseminations (IUI) combined with superovulation has made the assessment of sperm-cervical mucus interactions merely an academic exercise. In this view, whether the postcoital test is...
The iodine may exert a bacteriostatic effect on the mucous membranes. It may provide a stimulatory effect for the cilia of the tube. It may straighten the tubes and thus break down peritoneal adhesions.

Cavity by means of a #8 pediatric Foley catheter with the balloon inflated in the cervix, followed by ultrasonography of the uterus. Similarly, saline or albumin-based used, venous or lymphatic intravasation can be detected immediately, and injection of dye halted. A common approach is to demonstrate tubal patency with a water media; if it does occur, slow injection with pauses is helpful. If the tubes fill but dye droplets do not spill from the ends of the tubes, the uterus should be pushed up.

The dye should be injected slowly so that abnormalities of the uterine cavity are not missed. This is of special importance in diethylstilbestrol-exposed women, many of whom have abnormalities of uterine contour. Usually no more than 3 to 6 mL of dye are required to fill the uterus and tubes. If the patient complains of cramping, the injection of dye should be stopped for a few minutes and fluoroscopy temporarily discontinued. Spasm is rare with Ethiodol, an oil dye that is our preferred medium; if it does occur, slow injection with pauses is helpful. If the tubes fill but dye droplets do not spill from the ends of the tubes, the uterus should be pushed up in the abdomen by means of the tenaculum or suction cup. This puts the tubes on stretch and may help to release dye from the fimbriated ends. The droplets seen coming from the tube are the result of mixing of the oil dye and peritoneal fluid. On occasion, injection of dye into a hydrosalpinx will produce a similar pattern, and a delayed film to show loculation of dye is crucial in differentiating this condition from normal spill where the dye is distributed throughout the pelvis.

Evidence of distal tubal disease on HSG is usually confirmed when laparoscopy is performed. However, a normal study does not rule out pelvic pathology, especially immunologic factor. These findings, specifically immotile or absent sperm, also should prompt inquiry concerning use of vaginal lubricants. If there are no motile sperm in the postcoital test. When postcoital tests repeatedly reveal no sperm, only sperm shaking in place, only immotile sperm, or only a few motile sperm, despite good mucus, and when the mucus is repeated poorly timed, timed IUI, without the risks and expense of superovulation, is appropriate therapy for 3-4 cycles.

What constitutes a poor postcoital test in terms of sperm numbers is controversial. Intrauterine insemination has been reported to enhance the chances for pregnancy when there were 3 or fewer sperm/HPF in the postcoital test but not when there were 5 or more sperm/HPF. 52 We continue to support a minimum level of one motile progressive sperm/HPF as compatible with normality. The finding of no sperm, all dead sperm or a large proportion of shaking sperm suggests a possible immunologic factor. These findings, specifically immotile or absent sperm, also should prompt inquiry concerning use of vaginal lubricants. If there are no motile sperm in the postcoital test, the pH of the mucus should be determined.

**Hysterosalpingography**

A history of pelvic inflammatory disease, septic abortion, ruptured appendix, tubal surgery, or ectopic pregnancy alerts the physician to the possibility of tubal damage. Pelvic inflammatory disease is unquestionably the major contributor to tubal infertility and ectopic pregnancies. Westrom’s classic studies with laparoscopically confirmed pelvic inflammatory disease indicated that the incidence of subsequent tubal infertility is approximately 12% after one episode of pelvic infection, 23% after two episodes, and 54% after three episodes. 53 The risk of ectopic pregnancy is increased 6–7-fold after pelvic infection. Almost one-half of patients who are eventually found to have tubal pathology and/or pelvic adhesions, however, have no history of antecedent disease. Many of these women will have elevated anti-Chlamydia antibodies, suggestive of prior infection. There have been a few reports of damaged tubes showing histologic evidence of viral infection which could explain the absence of traditional causes of tubal damage.

Tubal disease is diagnosed by the hysterosalpingogram (HSG) and by laparoscopy. The HSG is performed 2 to 5 days after cessation of a menstrual flow. If there is a history suggestive of pelvic inflammatory disease, a sedimentation rate is obtained prior to the HSG and, if elevated, antibiotic therapy is given. The procedure is then postponed for 7 days when a repeat sedimentation rate is obtained. Only if this is normal is the HSG scheduled. If masses or tenderness are revealed by the pelvic examination at any time, the HSG should be bypassed and the pelvis evaluated by laparoscopy. If there is a documented history of pelvic inflammatory disease, the risk of a serious reinfection following HSG is too high, and it should be replaced by laparoscopy. If an HSG is performed in a patient who is at questionable risk for infection, the HSG should be performed with a non-PF oil dye rather than an oil dye that is normally used.

The majority of pregnancies that followed HSG occurred within 7 months of the procedure. A review of the question of oil versus aqueous dye noted that in every major study, the rate of pregnancy was higher with oil media; if it does occur, slow injection with pauses is helpful. If the tubes fill but dye droplets do not spill from the ends of the tubes, the uterus should be pushed up in the abdomen by means of the tenaculum or suction cup. This puts the tubes on stretch and may help to release dye from the fimbriated ends. The droplets seen coming from the tube are the result of mixing of the oil dye and peritoneal fluid. On occasion, injection of dye into a hydrosalpinx will produce a similar pattern, and a delayed film to show loculation of dye is crucial in differentiating this condition from normal spill where the dye is distributed throughout the pelvis.

In a randomized, prospective study of close to 400 women, the pregnancy and live birth rates were increased in women who had their HSG performed with oil dye. 54 Retrospective study in which increased pregnancy rates were noted after HSG, an oil dye was used. 55, 56 In a retrospective study where both tubes were patent on x-ray, the pregnancy rate was only slightly higher (58%) than when there was unilateral patency and nonfilling of the other tube (50%). 57 This finding has been confirmed in a more recent study. 58

Evidential tubal disease on HSG is usually confirmed when laparoscopy is performed. However, a normal study does not rule out pelvic pathology, especially adhesions or endometriosis. While the diagnostic usefulness of the HSG is established, its value as a therapeutic procedure in infertility is a subject of some controversy. Does an HSG enhance fertility, and, if so, is the effect seen with both oil and water-soluble dyes? A conception rate of 41.3% within 1 year of an HSG with oil media has been reported, whereas the rate was only 27.3% when water-soluble agents were employed. 59 This is in accord with other reports where the great majority of pregnancies that followed HSG occurred within 7 months of the procedure. The result also can suggest the need for intrauterine inseminations. Individuals with progressively motile sperm in the postcoital test have significantly higher pregnancy rates than those with nonmotile sperm.

In a randomized, prospective study of close to 400 women, the pregnancy and live birth rates were increased in women who had their HSG performed with oil dye. 60 Within 9 ovulation cycles following an HSG, the pregnancy rate was 33% in the oil dye group and 17% in the water dye group. However, another prospective, randomized study indicated no difference in subsequent pregnancy rates following HSG with oil or water dye. 61 Nevertheless, most of the evidence indicates that HSG with Ethiodol is a very useful therapeutic as well as diagnostic tool in women with infertility. 62

The effect of the oil dye on fertility could involve any of a number of mechanisms:

1. It may produce a mechanical lage of the tubes, dislodging mucus plugs.
2. It may straighten the tubes and thus break down peritoneal adhesions.
3. It may provide a stimulatory effect for the cilia of the tube.
4. It may improve the blood flow to the cervix.
5. The iodine may exert a bacteriostatic effect on the mucous membranes.
6. Ethiodol decreases in vitro phagocytosis by peritoneal macrophages. If the same effect occurs in vivo it could decrease macrophage activity and thus aid fertility by inhibiting the release of cytokines and decreasing phagocytosis of sperm. 63, 64

The use of an oil medium has been criticized on grounds that it is only slowly absorbed and may cause granuloma formation. Granulomas are found very infrequently, and they also can follow the use of water dyes. An additional fear with oil dye is embolization. However, only 13 cases of dye intravasation were encountered in 533 HSGs performed with Ethiodol. 65 Six of these women had embolization of the dye, but there were no symptoms, and no morbidity was noted. When fluoroscopy is used, the distinctive radiopaque intravasation can be detected immediately along with injection of dye halted. A common approach is to demonstrate tubal patency with a water soluble dye, then follow with Ethiodol for its fertility-enhancing effects.

HSG is less sensitive than sonohysterography in the identification of uterine polyps or myomas. Sonohysterography involves the injection of saline into the uterine cavity by means of a #8 pediatric Foley catheter with the balloon inflated in the cervix, followed by ultrasonography of the uterus. Similarly, saline or albumin-based fluids can be tracked by ultrasound as they course through the tubes into the peritoneal cavity, providing evidence of tubal patency without the use of x-ray. 66
Sonohysterography with a contrast solution of galactose microbubbles in galactose (Echovist) is being utilized in Europe to provide an overall evaluation of the pelvis, the uterus, and the fallopian tubes. 103, 104

Hysteroscopy

Hysteroscopy is a technique that complements hysterosalpingography. Direct visualization of the uterine cavity with a hystroscope is good for differentiating between endometrial polyps and submucous leiomyomas, establishing the definitive diagnosis and treatment of intrauterine adhesions, and for the diagnosis and treatment of intrauterine congenital anomalies. One can argue from a cost-effective point of view that hysterosalpingography and sonohysterography are more useful screening procedures. The hystroscope should be reserved to pursue abnormalities identified by other techniques, especially when operative intervention is planned.

Fallopocpy

Because of its narrow and tortuous character, it has been difficult to pass probes via the uterine cavity into the fallopian tube. This problem was overcome by the development of self-seeking guidewires and the adaptation of techniques used for coronary angioplasty. Hysteroscopic directed fallopocpy can be utilized to transvaginally examine the entire length of the tubal lumen. 105 This technique requires considerable expertise, but it has verified that the tubal ostium can undergo spasm, and intraluminal debris is present that can be a cause of tubal obstruction. The latter can be cleared by cannulation or balloon tuboplasty, or even by hysterosalpingography. Visualization of the inner aspect of the tube can also be accomplished from the distal end of the tube at the time of laparoscopy. Fallopocpy, originating at either end of the tube, can provide a good assessment of the tubal lining, indicating the prospect of success with tubal surgery.

Outpatient Canalization of the Tube

Proximal tubal obstruction can be treated by outpatient tubal cannulation or balloon tuboplasty. Transcervical tuboplasty can be performed by either a fluoroscopic or hysteroscopic approach, although most of the experience thus far is with the fluoroscopic technique. 106, 107 and 108 Discomfort can be minimized with intravenous sedation and a paracervical block. Cannulation and balloon tuboplasty success is achieved in at least one tube in 80–90% of attempts. Approximately 30% of patients will become pregnant in the 3–6 months following the procedure. The advantage of these accomplishments is the avoidance of expensive surgery or assisted reproductive technology.

Disorders of Ovulation

Disorders of ovulation account for approximately 20% of all infertility problems in couples. These can be anovulation or severe oligoovulation. In the latter cases, even though ovulation does occur, its relative infrequency decreases the woman’s chances for pregnancy. If periods occur only every 3 or 4 months, for practical purposes it matters little whether these are ovulatory or anovulatory. Anovulatory or oligoovulatory women should be promptly treated with clomiphene citrate to increase the frequency of, or to initiate, ovulation (see Chapter 30, and the drug can be started immediately, even before other areas have been investigated. If ovulation is the only infertility factor, most couples will become pregnant within 3 months of ovulation induction. Women with amenorrhea or hyperandrogenic anovulation should be evaluated and managed according to the clinical approaches detailed in Chapter 11, Chapter 12, and Chapter 13.

Basal Body Temperature

Women who have menstrual periods at monthly intervals marked by premenstrual symptoms and dysmenorrhea are almost always ovulatory, but not always; 5% are anovulatory. Indirect confirmatory evidence of ovulation can be obtained by use of basal body temperature (BBT) charts. The temperature can be taken orally with a regular thermometer or with special instruments (unnecessarily expensive however) that show a range of only a few degrees and thus are easier to read. The temperature is best taken immediately upon awakening and before any activity. The woman may be surprised to find that the basal temperatures are substantially lower than the usual 98.6°F (37°C). Days when intercourse takes place should be noted on the chart, and this may give the physician an indication that coital frequency is a problem.

Use of the BBT chart has been criticized because a small percentage of women who ovulate have monophasic graphs, and there is often disagreement concerning interpretation of individual charts. Moreover, the time of ovulation predicted by the BBT does not always correlate well with measurements of the LH surge or with perceptions of maximal cervical mucus production. There is a relationship between a nadir in the BBT and the LH surge, but the BBT is reliable in predicting the day of the LH surge only within 2–3 days. 109 Although the nadir is believed to represent the beginning of the LH surge, the occurrence of a nadir is variable and often is not detected. To be used prospectively to predict ovulation, nearly absolute cycle regularity is required, and this occurs infrequently.

Nevertheless, we still find the BBT helpful as a preliminary indicator of ovulation and as a tool for advising patients about the timing of intercourse. Patients should not become fixated on taking their temperatures, and usually several months of charts are sufficient.

![Graph showing basal body temperature chart](image)

A significant increase in temperature is not noted until 2 days after the LH peak, coinciding with a rise in peripheral levels of progesterone to greater than 4 ng/mL. 110 Physical release of the ovum probably occurs on the day prior to the first temperature elevation. The temperature elevation should be sustained for 11 to 16 days, falling at the time of the subsequent menstrual period.

If an approximate time of ovulation can be determined by temperature charts, a sensible schedule for coitus is every 36 to 48 hours in a period encompassed by 3 to 4 days prior to and 2 days after expected ovulation. It is unwise, however, to demand rigid adherence to a schedule. This may produce psychologic stress sufficient to inhibit sexual relations.

In discussing coital timing, the patients frequently want to know the fertilizable life of the sperm and the egg. The fertilizable life of the human oocyte is unknown, but most 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. 111 The great majority of pregnancies occur when coitus takes place within the 3-day interval just prior to ovulation. 112

Home urine LH testing is commonly used to assist diagnostic and therapeutic timing. The postcoital test should be performed within 12 hours of a positive urinary LH test. 113 However, appropriate use of the basal body temperature chart yields results equivalent to those obtained with the more expensive methods of urinary LH
A reliable assessment of ovulation can be obtained by endometrial biopsy. Endometrial biopsy is performed 2 to 3 days prior to the expected period, although a biopsy performed in the midluteal phase is superior for diagnosing luteal phase defects. 111 The histology is read by the criteria outlined by Noyes, Hertig, and Rock. 112 Although premenstrual biopsy could interrupt a pregnancy if performed in a conception cycle, the danger is not great. 113 We recommend the use of the plastic endometrial suction curette. It is easy to use, requires cervical dilation only occasionally, and is usually painless.

**Progestrone Measurements**

A serum progesterone level of less than 3 ng/mL is consistent with follicular phase levels. 112 To confirm ovulation, values at the midluteal phase, just at the midpoint between ovulation and the onset of the subsequent menstrual period, should be at least 6.5 ng/mL and preferably 10 ng/mL or more. The consensus of opinion is that a single midluteal phase progesterone level is insufficient evidence upon which to judge the adequacy of the luteal phase. 112, 114, 116, 119, 121 and 113. The progesterone level is subject to the variation associated with pulsatile secretion, but more importantly, there is often poor correlation with the histologic state of the endometrium.

**Luteal Phase Defect (Inadequate Luteal Phase)**

A luteal phase defect, historically defined as a lag of more than two days in histologic development of the endometrium compared to day of the cycle (presumably due to inadequate progesterone secretion or action), can be found in up to 30% of isolated cycles of normal women, and only if the defect is found repeatedly is it thought to be a possible factor in infertility. Approximately 3 to 4% of infertile women will be diagnosed as having luteal phase defect, and the incidence may be higher (approximately 5%) in women with a history of recurrent abortions. 112, 117

Although luteal phase defect is often a direct result of decreased hormone production by the corpus luteum, the underlying causes of this dysfunction can be multiple. Decreased levels of FSH in the follicular phase of the cycle, abnormal patterns of LH secretion, decreased levels of LH and FSH at the time of the ovulatory surge, or decreased response of the endometrium to progesterone have been implicated. 117 Elevated prolactin levels and hypothyroidism can be associated with luteal phase defect, emphasizing the importance of assessing prolactin and TSH levels in infertile women with less than normal ovulatory cycles.

The diagnosis should be considered in women with normal cycles and unexplained infertility, women with short luteal phases demonstrated by basal body temperature charts, and women with a history of recurrent spontaneous abortions.

Controversies surrounding the concept of luteal phase defect have revolved around issues of diagnosis, endometrial biopsy versus progesterone levels, and treatment, progesterone versus clomiphene citrate. However, a more fundamental question must be addressed. Is there really such an entity as luteal phase defect, and even if there is, does it play any role in infertility?

As noted above, although measuring progesterone levels has been advocated as a means of diagnosing luteal phase defect, the majority of clinical studies on this subject have used endometrial biopsy as the standard. The endometrium must lag behind the day of the cycle by greater than two days, and the lag must occur in more than one cycle. Whereas it is common to date the cycle day from the onset of the subsequent menstrual period, there is evidence that better dating can be achieved by counting forward from the LH surge. 118, 119

An important issue is how well can the physician, usually a pathologist, date the endometrium. When the same observer viewed 63 endometrial biopsies on two separate occasions, there was exact agreement in only 15 (24%). 120 Disagreement of more than two days in the dating of the endometrium occurred in six instances (10%). A further problem in establishing luteal phase defect as a clinical entity is the finding in most studies of an out-of-phase biopsy in approximately 20% to 30% of normal cycles and repetitive lags in more than one normal cycle of approximately 5%. 121, 122 These figures suggest that luteal phase defect occurs by chance alone.

The frequency of luteal phase defect in women with infertility, when strictly defined, is no greater than that found by chance in normal cycles. 117 In a series of 1482 biopsies in 1555 women reported by Balasch, there were 26 biopsies in conception cycles. 123 With an in-phase biopsy, 15 of 20 pregnancies went to term but 4 out of 6 pregnancies with out-of-phase biopsies also went to term. Similarly, the term pregnancy rates were almost identical in women treated and untreated for luteal phase defect. There is a window of endometrial hormonally-induced development during which successful embryo transfer with in vitro fertilization can occur. 125 The window extends over a 6-day period, and this indicates that precise synchronization of the endometrium and the embryo may not be needed for successful implantation.

Given the uncertainties surrounding the diagnosis of luteal phase defect, it becomes difficult to justify putting women through the expense and possible pain of endometrial biopsy.

In common practice, again because of the discomfort, the difficulties of precise dating, and the expense associated with endometrial biopsy, attention has turned to measurements of serum progesterone levels as a means of diagnosing, if not luteal phase defect, then at least a “hormone deficiency.” Whereas exact normal values for progesterone are in some dispute, many physicians believe that a level of less than 10 ng/mL one week prior to the onset of menstruation or 5–7 days after the LH surge is a good indication of a luteal phase defect. Frequently a diagnosis of “hormone deficiency” is made based on an isolated and not always well timed progesterone level of less than 10 ng/mL. It is also common practice to utilize such a finding as a rationale for treating as if a luteal phase defect were present. Daily progesterone measurements taken throughout the luteal phase could provide strong evidence for a luteal phase defect if the values are low, but such frequent sampling is impractical. An assessment of various approaches for progesterone measurements concluded that the best correlation with endometrial histology is achieved with 3 progesterone levels obtained randomly during days 5–9 after ovulation (normal equal to or greater than a sum of 30 ng/mL or a pooled concentration, a less expensive alternative, greater than 9 ng/mL). 127 Most important, however, is the impressive evidence documenting poor correlation between progesterone measurements and endometrial histology. 128, 129

Other methods of diagnosis include the basal body temperature chart, ultrasonography of follicular size, and salivary progesterone measurements. All have been studied, and all demonstrate poor correlations with multiple progesterone levels and endometrial histology. 128, 130
Treatment of Luteal Phase Defect

Based on findings that low FSH values prior to ovulation can be associated with luteal phase defect, it would seem reasonable, in selected cases, to use human gonadotropins or clomiphene citrate. Gonadotropin treatment is expensive, has the potential for causing hyperstimulation of the ovaries, and is associated with an increased incidence of multiple births. Because of these effects, gonadotropin treatment is seldom, if ever, used for this indication. Clomiphene citrate is the first choice of many clinicians for the treatment of luteal phase defect. The only significant risk is a 5% (twice normal) chance of multiple births, essentially all twins. The initial dose is 50 mg a day for 5 days starting on day 3, 4, or 5 of the cycle (Chapter 30).

Because there is a suspected deficiency of progesterone in luteal phase defect, exogenous progesterone has been utilized. A vaginal suppository compounded by a pharmacist and containing 25 mg of progesterone is inserted bid starting approximately 2–3 days after ovulation. Treatment is maintained until menstruation occurs or through the 10th week of a pregnancy. Once pregnancy is diagnosed, a switch can be made to weekly injections of 17-hydroxyprogesterone caproate (250 mg) through the 10th week of pregnancy. Oral micronized progesterone can be utilized in a dose of 100 mg tid. However, the oral administration of progesterone leads to metabolic products that cause somnolence and other CNS side effects. Progesterone in a gel that adheres to the vaginal mucosa (Crinone) is available in a prefilled vaginal applicator, applied with one 8% (90 mg progesterone) application daily. Vaginal administration accomplishes targeted delivery to the uterus without producing high circulating levels. Using progesterone therapy, success rates of approximately 50% have been achieved, but good controlled studies are lacking.

There is no difference in pregnancy rates in studies comparing clomiphene and progesterone treatment. In our view, this is an argument in favor of clomiphene because of a significant disadvantage associated with progesterone therapy. Progesterone supplementation prolongs the luteal phase and can delay the onset of menses. This is not a problem for the clinician, but for the couple the disappointment at the time of delayed menses or a negative pregnancy test is profound.

Dopamine agonist treatment has been reported to correct luteal phase defect associated with hyperprolactinemia, but its value in women with normal prolactin levels has not been demonstrated. In a subgroup of patients with unexplained infertility, high normal prolactin levels, and expressible galactorrhea, treatment with bromocriptine enhanced fertility compared to similar women treated with pyridoxine. If galactorrhea is present, even if the prolactin is normal, ovulatory dysfunction responds well to dopamine agonist therapy. In the absence of galactorrhea, a prolactin elevation may be subtle (such as an increase in nocturnal peaks), and this could explain occasional good responses to dopamine agonist treatment. In evaluating any therapy it is important to keep in mind that pregnancies can occur without treatment in women who are diagnosed as having luteal phase defect.

Many physicians dispense with the diagnostic evaluation of hormone adequacy and proceed to treatment with clomiphene citrate. It is a reasonable argument that there may be subtle hormonal abnormalities that cannot be diagnosed with current technology but that can be successfully treated by stimulating the ovaries. Moreover, there is a theoretical advantage in having more than one oocyte in the fallopian tube at the time of fertilization. Randomized placebo-controlled studies support the efficacy of this approach (reviewed with references in Chapter 30). By contrast, some have not found a benefit for the use of clomiphene citrate in unexplained infertility.

We do not object to the empirical use of clomiphene, provided there is a clear understanding of the potential side effects and the uncertain efficacy of the medication.

Luteinized Unruptured Follicle

On occasion, a corpus luteum will form despite the failure of release of the oocyte. Initially it was thought that this problem could be identified at laparoscopy by visualizing an absence of the ovulatory stigma, but now it is apparent that the stigma can be epithelialized rapidly and thus obscured from view. Currently, clinical diagnosis of a luteinized unruptured follicle (LUF) is made on the basis of ultrasound monitoring. The preovulatory growth of the follicle usually is normal but the follicle does not collapse following the LH surge, and there may be increased growth in the luteal phase. The interior of the follicle lacks the echoes often seen in corpora lutea. Whereas these criteria seem straightforward, establishing the diagnosis of LUF is often difficult. Even if ultrasonography is performed daily, the collapse of the follicle can be missed, and a corpus luteum refilled with blood can be mistaken for a persistent follicle. Therefore, serial ultrasound screening of women with unexplained infertility is of questionable value, not only for the diagnosis of LUF, but also for determining the size of the follicle just prior to rupture. Whereas ovulation usually occurs in the normal cycle when the follicle is 17–23 mm in diameter, some fertile women ovulate from a smaller follicle. In the evaluation of unexplained infertility, a single ultrasound examination in the late follicular phase is of value to detect endometrial polyps and myomas.

It is doubtful that LUF is a significant cause of infertility, and furthermore, the suggested treatments are supereovulation, either alone or in conjunction with one of the assisted reproductive technologies, treatment choices that will be empirically offered even without a diagnosis of LUF. Because inhibition of prostaglandin synthesis can cause a luteinized unruptured follicle, patients should be cautioned to avoid the use of nonsteroidal antiinflammatory agents.

Mycoplasma

Mycoplasma, a pleuropneumonia-like organism, has been implicated as a possible cause of recurrent abortion and salpingitis. A markedly higher prevalence of T ureaplasma (ureaplasma urealyticum) was detected in the cervical mucus and semen of infertile couples compared with a group of fertile couples. Treatment with doxycycline decreased the number of couples with mycoplasma and also was associated with pregnancy in 15 of 52 couples (29%), all of whom had had primary infertility of at least 5 years duration. However, a series of reports from England agreed with these findings in only one respect. They confirmed that treatment with doxycycline could eliminate mycoplasma from the genital tract of the majority of individuals. There was no difference, however, in the frequency of either T strain or Mycoplasma hominis between infertile and fertile couples. In a double-blind study, treatment with doxycycline for 28 days had no effect on the rate of conception, and a English group suggested that culturing for mycoplasma in the routine investigation of infertility was unrewarding.

Since those early publications, a number of studies have established the widespread distribution of ureaplasma urealyticum in both fertile and infertile populations. Some have found higher colonization in infertile couples, whereas others have found no relationship between the organisms and infertility. In a study that received a great deal of media attention, it was reported that 60% of males who were culture positive for ureaplasma urealyticum and were cleared of infection by antibiotic treatment achieved a pregnancy. Failure to clear the infection resulted in a 5% pregnant rate. This study suffered from lack of clarity on the criteria for entry into treatment and from any mention of individuals lost to follow-up. The incidence of ureaplasmal infection is only significantly higher in those women whose male partners have semen abnormalities.

Ureaplasma urealyticum is widespread, but there is no compelling evidence that it causes infertility. Empiric treatment with antibiotics is not warranted.

Endoscopy

Traditionally, laparoscopy has been the final diagnostic procedure of any infertility investigation. As the success rates with in vitro fertilization improve (Chapter 31), patients and physicians increasingly believe that when diagnostic laparoscopy is performed by an experienced surgeon who offers laparoscopic technologies is appropriate, even without laparoscopy. In most hands, in vitro fertilization will be more successful than treating significant tubal disease through the laparoscope. Mild, asymptomatic endometriosis of prolonged duration may also require in vitro fertilization if supereovulation, combined with intrauterine inseminations, does not achieve pregnancy. Considerations of cost and healthcare coverage often influence the types of treatment selected by couples.

If laparoscopy is contemplated after a normal HSG, the endoscopic procedure is usually performed after an interval of 6 months from the x-ray. This allows time for the fertility enhancing effect of the x-ray procedure. An exception can be made for the woman who is at high risk for pelvic infection or the older woman who has no time to wait. The findings at laparoscopy agree with those of HSG in approximately two-thirds of the cases. The major area of disagreement is the failure of the HSG to detect pelvic adhesions or endometriosis. Approximately 50% of apparently normal infertile patients undergoing laparoscopy will have pelvic pathology, usually pelvic adhesions or endometriosis (but ordinarily a degree of endometriosis not requiring treatment). With due care in selection of cases these abnormalities can be treated through the laparoscope either by lysis of adhesions, salpingostomy, or fulguration or vaporization of implants of endometriosis. Patients with significant tubal disease are best advised to proceed to in vitro fertilization.
Unexplained Infertility

An infertile couple has what is called unexplained infertility when all standard clinical investigations (semen analysis, the postcoital test, assessment of ovulation, demonstration of tubal patency) yield normal results. It is estimated that from 10% to 15% of infertile couples will ultimately reach this clinical diagnosis, and, using normal findings on laparoscopy as a criterion, the prevalence may be less than 10%. Important variables are age of the woman and duration of the infertility.

The average monthly fecundity in normal couples is 30%; the monthly pregnancy rate in couples with unexplained infertility is 1.5–3%. After 3 years of infertility, the prospect of pregnancy decreases by 24% each year. Approximately 80% of couples with unexplained infertility of less than 3 years duration will become pregnant with 3 years of expectant management. Because the incidence of spontaneous pregnancy is significant until 3 years have passed, it is appropriate to require 3 years of infertility in women less than 30 years old before making this diagnosis. Because age is an important determinant, further evaluation and therapy should not be deferred in older women.

A meticulous review of available results is essential to avoid overlooking a treatable factor. The use of sperm function tests may be helpful. There is some correlation between absent sperm penetration of hamster eggs and subsequent outcome (see Chapter 29). If these tests are not available, keep in mind that a definite diagnosis of unexplained infertility requires successful fertilization in vitro. Thus, a human egg test (in vitro fertilization) is the ultimate test of sperm fertility, and it is more time and cost efficient to move promptly to assisted reproduction, rather than testing sperm function.

Superovulation and Intrauterine Insemination

Empiric treatment for suspected endometriosis or with dopamine agonists has no impact on unexplained infertility. However, the methods of assisted reproductive technology and superovulation with intrauterine insemination do increase the prospect of pregnancy. The results with superovulation alone are inferior to those achieved with one of the assisted reproductive techniques, but the overall cost is also less. Studies indicate that subtle hormonal abnormalities do exist in women with unexplained infertility. Changes (elevated gonadotropins and estradiol in the follicular phase and decreased luteal phase progesterone) have been attributed to a reduction in ovarian reserve as usually seen in older women. Thus, there is a rational basis to provide ovulation stimulation.

A cumulative pregnancy rate of 40% can be achieved in couples with unexplained infertility after 6 cycles of superovulation or 3 cycles of in vitro fertilization. In randomized, controlled clinical trials, the monthly pregnancy rates in couples with unexplained infertility is increased 2–3-fold (a monthly fecundity rate of 9%) with clomiphene treatment, and with human gonadotropins, the monthly fecundity rate is approximately 10–15%.

Although some studies have failed to yield increased pregnancy rates with the empiric use of clomiphene, we believe couples with unexplained infertility should be offered superovulation or one of the assisted reproductive technologies. However, superovulation should be limited to 3 or 4 cycles.

The addition of intrauterine insemination of washed sperm (IUI) to clomiphene treatment increased fecundity in couples with unexplained infertility or surgically treated endometriosis, compared with intercourse without hormonal stimulation. As with almost all infertility, success is age-related, with pregnancy rates 3–4-fold greater in women under age 20. Although the combination of IUI with clomiphene may enhance the pregnancy rate in couples with unexplained infertility, IUI by itself does not.

The combination of IUI with gonadotropin stimulation provides even better results than clomiphene and IUI. In two studies, the pregnancy rate per cycle with gonadotropins and IUI was 19% and 14.3% compared to 4% and 7.7% with clomiphene and IUI respectively.

In a meta-analysis of randomized trials comparing gonadotropin stimulation with IUI or with timed intercourse, IUI was associated with approximately a 2-fold greater pregnancy rate. In a randomized trial, superovulation and IUI resulted in a 3-fold greater pregnancy rate compared with intracervical insemination and 2-fold greater compared with either IUI alone or a combination of superovulation and intracervical insemination.

The effectiveness of gonadotropin treatment combined with IUI has been repeatedly demonstrated for a variety of infertility disorders, including endometriosis, cervical factor, and unexplained infertility. As with clomiphene, few pregnancies occur after 4 cycles of treatment.

Thus, there exist a number of options for the empiric treatment of unexplained infertility before considering the more complex and costly assisted reproductive technologies (Chapter 31). These options and the timing of their use should always be considered in the context of the known spontaneous cure rate for unexplained infertility. Approximately 60% of couples with unexplained infertility of less than 3 years duration will become pregnant within 3 years of expectant management. However, as time progresses, the monthly fecundity rate decreases, and this can be improved by active management with superovulation and IUI. For older women, decreasing the time to achieve a pregnancy is an important objective.

In a meta-analysis of methods used to treat unexplained infertility, pregnancy rates were estimated to be as follows:

<table>
<thead>
<tr>
<th>Method</th>
<th>Pregnancy Rate per Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>No treatment</td>
<td>1.3–4.1%</td>
</tr>
<tr>
<td>IUI</td>
<td>3.8%</td>
</tr>
<tr>
<td>Clomiphene</td>
<td>5.6%</td>
</tr>
<tr>
<td>Clomiphene &amp; IUI</td>
<td>8.3%</td>
</tr>
<tr>
<td>Gonadotropins</td>
<td>7.7%</td>
</tr>
<tr>
<td>Gonadotropins &amp; IUI</td>
<td>17.1%</td>
</tr>
<tr>
<td>IVF</td>
<td>20.7%</td>
</tr>
</tbody>
</table>

Because it is relatively inexpensive, easy to administer, and not usually burdened with major side effects (except for a 2-fold increase in twinning), clomiphene is usually the first treatment choice. Patient demand for this treatment is substantial, and it is not unreasonable to respond to that demand with 3–4 clomiphene cycles. The uncertainty regarding the efficacy of adding IUI to clomiphene influences us to proceed from clomiphene treatment to the combination of gonadotropins and IUI, with an expectation of achieving a pregnancy rate of 15% per cycle. The combination of clomiphene and IUI is still appropriate in circumstances where cost, time, and avoidance of complications are paramount. The use of IUI alone or gonadotropins without IUI is not recommended for the treatment of unexplained infertility.

The success of treatment is influenced by the number of sperm inseminated. Pregnancies are rare if less than 1–3 million sperm are inseminated. Once this baseline is surpassed, increasing the number of sperm inseminated does not increase the pregnancy rate; although one study achieved higher pregnancy rates with inseminations of greater than 10 million motile sperm. Morphology plays only a minor role. Once the percentage of normal forms exceeds 5–10%, there is no influence on success.

The effects of age remain critical whatever the method of ovarian stimulation. In one series of 136 treatment cycles, no pregnancies were achieved in women who were 43 or older. Despite rare successes in older women, it borders on the exploitive to treat women over age 40 with superovulation and IUI. Indeed, pregnancy rates begin to decline with this treatment around age 35.
Adoption

With proper assessment and therapy, the majority of couples evaluated for infertility will become pregnant. For those who are refractory to the usual treatments, consideration of assisted reproductive technologies or adoption is appropriate. Couples thinking about adoption have a range of choices including social agency adoption, and international adoption. In some states, traditional adoption is not legal; however, it can provide an effective, more rapid alternative to adoption through a social agency. In most cases the biological mother has the opportunity to know the adopting parents, and this lack of anonymity may direct some individuals away from private adoption. In addition, there is a short time period during which the biological mother can reclaim the baby. In our experience, this devastating event occurs in approximately 5% of private adoptions.

To facilitate private adoption, patients should be encouraged to "spread the word" that they are interested in adoption. In addition, letters can be directed to obstetricians throughout the country describing the couple and their desire for adoption. Both lawyers and non-lawyers have practices devoted to counseling couples (and individuals) on approaches to private adoption and regarding the adoption laws in individual states.

Myths and Appropriate Goals

It is important for physicians and other health care professionals to dispel the myths that are associated with infertility. Women should not be told that they are infertile because they are too nervous. Unless anxiety interferes with ovulation or coital frequency, there is no present evidence that infertility is caused by the usual anxieties. It is important for physicians and other health care professionals to dispel the myths that are associated with infertility. Women should not be told that they are infertile because they are too nervous. Unless anxiety interferes with ovulation or coital frequency, there is no present evidence that infertility is caused by the usual anxieties.

Chapter References

Early pregnancy loss (abortion or miscarriage) is defined as the termination of pregnancy before 20 weeks of gestation (dated from the last menstrual period) or below a fetal weight of 500 g. Approximately 15% of all pregnancies between 4–20 weeks of gestation will undergo clinically recognized spontaneous miscarriages. The true early pregnancy loss rate is closer to 50% because of the high rate of unrecognized miscarriages in the 2–4 weeks immediately following conception. The majority of these very early cases are caused by chromosomal abnormalities in the sperm or the egg.

“Habitual” abortion was classically defined as 3 or more consecutive spontaneous miscarriages. In 1938, Malpas, using theoretical calculations, stated that a woman with a history of 3 consecutive spontaneous miscarriages had a 73% chance of aborting in the next pregnancy. In 1946, Eastman presented statistical calculations indicating that after 3 miscarriages the risk was 83.6%. These early conclusions, based primarily upon intuition rather than clinical studies, established the notion that the chance for a subsequent miscarriage increases dramatically with each successive miscarriage; and that after 3 miscarriages the chances for a successful pregnancy are very low. Studies on the efficacy of many types of treatments used these pessimistic figures for comparison rather than containing their own controls. If treatment increased the salvage rate to 70% it was considered curative. However, clinical studies have indicated that the risk of pregnancy loss after 3 consecutive spontaneous miscarriages is in fact 30–45%. The chance of a successful live birth after 3 consecutive miscarriages without a live birth is 55–60%; with at least one previous normal pregnancy (live birth) in addition to the repetitive miscarriages, the chance is approximately 70%. However, keep in mind that these numbers are derived from younger women, and older women have a risk of pregnancy loss that is twice that of younger women.

<table>
<thead>
<tr>
<th>The Risk of Recurrent Early Pregnancy Loss in Young Women</th>
<th>Number of Prior Miscarriages</th>
<th>% Risk of Miscarriage in Next Pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women who have had at least one liveborn infant:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>12%</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>24%</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>26%</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>32%</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>26%</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>53%</td>
</tr>
<tr>
<td>Women who have not had at least one liveborn infant:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or more</td>
<td></td>
<td>40–45%</td>
</tr>
</tbody>
</table>

The projections by Malpas and Eastman were theoretical exercises that were not confirmed when appropriate data were collected. Thus, it is not surprising that treatment with a wide range of approaches, including vitamins and psychotherapy, produced successful pregnancies in a reasonable percentage of women with recurrent miscarriages. These cures were not due to the therapy; the claims for success were based on a comparison with the discredited statistics of Malpas and Eastman.

The diagnostic and therapeutic response to a couple with pregnancy loss is not dictated by the number of miscarriages. The response is significantly influenced by the woman's age, the couple's level of anxiety, and factors readily identified in the family and medical history. The degree of response will range from an educational discussion to a full diagnostic evaluation with appropriate treatment. It is helpful to consider recurrent pregnancy losses according to the following categories:

1. Normal statistics,
2. Genetic factors,
3. Environmental factors,
4. Endocrine factors,
5. Anatomic causes,
6. Infectious causes,
7. Thrombophilia,
8. Immunologic problems.

Normal Statistics

The reproductive loss between conception and clinically recognizable pregnancy is significant; about 50% of fertilized ova do not progress to a viable pregnancy. The use of sensitive assays for human chorionic gonadotropin (hCG) suggests that up to 30% of pregnancies are lost between implantation and the 6th week. It is important for physicians and their patients to be aware of the high degree of reproductive loss, especially in older women due in part to the increasing frequency of trisomies with advancing age. However, the frequency of both euploid (normal) and aneuploid (abnormal) abortuses increases with maternal age.

Approximately 80% of spontaneous miscarriages occur in the first 12 weeks of pregnancy, and nearly 70% of these miscarriages in early pregnancy are due to chromosomal anomalies. Clinically recognized miscarriage occurs in only 12% of women younger than age 20, but the incidence increases to 26% in women older than age 40. The overall rate of loss (recognized and unrecognized) in women over age 40 is approximately 75%. An appreciation for these statistics contributes significantly to a couple's ability to cope with an early pregnancy loss.

Once a live embryo is detected by ultrasonography in normal young women or in young women with infertility, the rate of fetal loss is 3–5%. However, in women with recurrent pregnancy loss, the rate of loss after detection of fetal cardiac activity is 4–5 times higher. The rate of spontaneous miscarriage in women with two or more recurrent losses was 22.7% after ultrasound documentation of fetal cardiac activity. Keep in mind that the risk of spontaneous miscarriage is higher in older woman; spontaneous miscarriage occurred in 20% of women 40 or more years old, undergoing in vitro fertilization, after the demonstration of fetal heart motion by ultrasonography.

Genetic Factors

Despite the knowledge that the spontaneous success rate is 55–70%, it is still worth trying to uncover causes for repetitive first-trimester miscarriages. A recognized cause of the problem is a genetic abnormality, and karyotyping of couples will reveal that 3–8% have some abnormality, most frequently a balanced chromosomal...
derived from randomized clinical trials. The septate uterus is the most frequent anatomic abnormality associated with recurrent early spontaneous miscarriages, and often by hysteroscopy, is rewarded with delivery rates in the 70–80% range; however, this high rate of success must be tempered by the realization that it is not of women with recurrent miscarriages have a uterine malformation, and this can be best diagnosed by vaginal ultrasonography (especially with saline instillation), and Uterine abnormalities can result in impaired vascularization of a pregnancy and limited space for a fetus due to distortion of the uterine cavity. Approximately 12–15% Anatomical Causes

recurrent miscarriage, even without treatment.

As noted, approximately 70% of early spontaneous miscarriages are associated with fetal chromosomal abnormalities. In addition, 30% of second-trimester miscarriages and 3% of stillbirths have abnormal chromosomes. In most cases, the couple is chromosomally normal and the fetal chromosomal abnormality is a random event. The abnormalities include maternal and paternal accidents in gametogenesis, as well as miscalculation. The fetal chromosomal abnormalities in single spontaneous miscarriages are different than those in recurrent miscarriages. Autosomal trisomy is the most frequent anomaly (about 50% of early spontaneous miscarriages), due to nondisjunction or translocations are assays of chromosomes 13, 16, 18, 21, and 22 are the most common. The next most common anomaly (about 25%) is 45,X which is responsible for Turner syndrome when the fetus survives. Of the remaining anomalies, most are polyploidy.

Fetuses that abort later in gestation usually have normal chromosomes. Perhaps in this situation, the responsible factors are external to the fetus. Subtle chromosomal defects, however, may be revealed by our analytic techniques improve.

According to McDonough, treatment of endocrine factors yields a 90% normal child rate; correction of anatomic factors yields a 60–70% rate, but known genetic factors are associated with only a 32% expectation for a normal child.

It is helpful to have a karyotype on a previous miscarriage to determine aneuploidy or euploidy. Once determined, there is an increased likelihood that subsequent miscarriages will be the same; although there is still a chance for women with recurrent pregnancy loss to have a normal pregnancy. If aneuploidy is documented, accurately timed inseminations could be considered based on animal studies relating aneuploidy to age of ovum and sperm; otherwise the choice is between hoping for the best or donor insemination.

Karyotyping is expensive. A factor that can help in decision making is a positive family history for recurrent miscarriages. Translocations within families are usually associated with a mixed history: some normal pregnancies intermixed with recurrent miscarriages. Karyotyping is indicated when family members can be identified with multiple spontaneous miscarriages or the family has a malformed or mentally retarded child or a child with a known chromosomal abnormality. In any event, we recommend a karyotype when there is a history of 3 consecutive spontaneous early pregnancy losses and no previous normal liveborn.

Environmental Factors

Smoking, alcohol, and heavy coffee consumption have been reported to be associated with an increased risk of recurrent pregnancy losses. The increase in risk is proportional to the number of cigarettes smoked. In these cases, the fetal chromosomes are normal. More recently, the link with caffeine intake could not be supported. Anesthetic gases and tetrachloroethylene (used in dry cleaning) have been implicated as teratogenic agents of miscarriage, but exposure to video terminals is not a factor. Exercise programs do not increase the risk of spontaneous miscarriage, and bed rest will not influence the risk of recurrent miscarriage. Isotretinoin (Accutane) is definitely associated with an increased incidence of spontaneous miscarriage. An increased risk of spontaneous miscarriage has not been documented among laboratory workers or women in the pharmaceutical industry. The use of electric blankets and heated water beds is also not associated with an increased risk of spontaneous miscarriage.

Endocrine Factors

Mild or subclinical endocrine diseases are not causes of recurrent miscarriages. Patients who have significant thyroid disease or uncontrolled diabetes mellitus may suffer spontaneous miscarriages, but it is unlikely that laboratory assessments of thyroid function and carbohydrate metabolism are worthwhile in relatively healthy women. Nevertheless, the high frequency of hypothyroidism warrants screening with a measurement of thyrotropin-stimulating hormone (TSH).

No convincing evidence exists linking endometriosis with an increased risk of spontaneous miscarriage.

Elevated luteinizing hormone (LH) levels have been associated with an increased risk of miscarriage in women with polycystic ovaries who become pregnant. There is no reason to believe that this is a factor in ovulatory women with recurrent miscarriages. Even in women with polycystic ovaries and a history of 3 or more consecutive early pregnancy losses, suppression of LH secretion with a GnRH agonist prior to ovulation induction yielded no difference in outcome.

An endocrine abnormality that may cause recurrent miscarriages is the inadequate luteal phase. Studies of the role of hormone deficiency as a cause of recurrent miscarriage have largely focused on deficiencies of progesterone or its metabolites. Attempts to implicate low progesterone or pregnanediol levels in early pregnancy as a cause for miscarriage, and, as a corollary, to treat with exogenous progesterone or progestins have been fruitless.

Another approach has been to diagnose an inadequate luteal phase during the nonpregnant state and to initiate treatment with progesterone a few days after ovulation. Jones and Defs claimed that 30% of women with pregnancy wastage had an inadequate luteal phase, whereas more recent studies have found that 20–25% of women with recurrent miscarriages have an inadequate luteal phase, Yet it is by no means certain that gestational treatment makes a difference. Indeed, a meta-analysis of randomized trials of pregnancies treated with gestational agents failed to find any evidence for a positive effect on the maintenance of pregnancy. Nevertheless, clinicians continue to report, as we have noted, a significant incidence of luteal phase inadequacy in their series of patients with recurrent miscarriages.

The uncertainty that plagues physicians in dealing with the inadequate luteal phase in infertility, therefore, is also apparent when considering recurrent miscarriages. If repetitive endometrial biopsies indicate a lag in histologic development of more than 2 days or if the basal body temperature chart shows a luteal phase of less than 11 days, it is reasonable to treat with clomiphene or progesterone (Chapter 20) and with a dopamine agonist if galactorrhea or an elevated prolactin is present. In the absence of galactorrhea or hyperprolactinemia, we prefer clomiphene because this avoids the situation of a prolonged luteal phase due to progesterone treatment, a false expectation of pregnancy, and a more difficult time for the couple in coping with the results of a negative pregnancy test.

Should clomiphene be offered empirically (for 6 cycles) when no other cause for recurrent miscarriages can be identified? Placebo treatment may be useful in maintaining a clinician-patient relationship in which the patient derives needed psychologic sustenance. Furthermore, there is no evidence of any major harmful consequences with clomiphene treatment (see Chapter 30). In view of the difficulty in establishing the diagnosis of an inadequate luteal phase and the uncertainties that surround this diagnosis, we believe there is a place for empirical treatment. It must be remembered, however, that there is a reasonable cure rate in unexplained recurrent miscarriage, even without treatment.

Anatomical Causes

Uterine abnormalities can result in impaired vascularization of a pregnancy and limited space for a fetus due to distortion of the uterine cavity. Approximately 12–15% of women with recurrent miscarriages have a uterine malformation, and this can be best diagnosed by vaginal ultrasonography (especially with saline instillation), and confirmed by magnetic resonance imaging. Hysterosalpingography is relatively inaccurate and decisions should not be based upon hysterosalpingography alone. The various uterine anomalies, including leiomyomata and diethylstilbestrol (DES) exposure, are discussed in detail in Chapter 4. Surgical repair of these defects, often by hysteroscopy, is rewarded with delivery rates in the 70–80% range; however, this high rate of success must be tempered by the realization that it is not derived from randomized clinical trials. The septate uterus is the most frequent anatomical abnormality associated with recurrent early spontaneous miscarriages, and
Our preferred treatment for significant titers of antiphospholipid antibodies consists of the combination of low-dose aspirin (80 mg daily) and low-dose heparin as soon as helpful to measure antinuclear antibodies, IGA antibodies, or leukocyte antibodies. Despite activating thrombosis, the antiphospholipid antibodies prolong the prothrombin time and the partial thromboplastin time. The activated partial thromboplastin time retardation and fetal death in addition to recurrent miscarriages, and when present, there is a high rate of second-trimester fetal deaths. The mechanism of pregnancy loss is probably decidual and placental insufficiency due to the thrombotic tendency. Annexin-V is a phospholipid-binding protein that inhibits coagulation; the levels of annexin-V on trophoblasts and endothelial cells are reduced in the presence of antiphospholipid antibodies.

Infectious Causes

Despite periodic reports that have implicated specific infectious agents as etiologic factors in recurrent miscarriages, there currently is no hard evidence that bacterial or viral infections cause recurrent pregnancy losses. An impressive incidence of antichlamydial antibody has been reported in women with 3 or more spontaneous miscarriages, but it is not certain whether this is an association with Chlamydia trachomatis or whether this is a marker of a different immune response in women with recurrent miscarriages. In a prospective study, no association could be detected between antichlamydia antibodies and spontaneous miscarriage.

Claims of effective antibiotic treatment have been derived from randomized studies. Perhaps an exception is infection with Ureaplasma urealyticum. Other organisms that have been implicated, but not substantiated, include Toxoplasma gondii, Listeria monocytogenes, Mycoplasma hominis, herpes virus, and cytomegalovirus. It is more cost-effective and time efficient to prescribe both partners a course of doxycycline (100 mg bid for 14 days) or erythromycin (250 mg qid for 14 days) than to pursue multiple and repeated cultures.

Thrombophilia

The major cause of thrombosis in pregnancy is an inherited predisposition for clotting, especially the factor V Leiden mutation. Deficiencies of antithrombin III, protein C, and protein S are inherited in an autosomal dominant pattern, accounting for 10–15% of familial thrombosis. Mutations in the prothrombin gene and the factor V Leiden mutation are the most common inherited causes of venous thromboembolism. 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. The factor V Leiden mutation is found in approximately 30% of individuals who develop venous thromboembolism. Activated protein C inhibits coagulation by degrading factors V and VIII. One of the 3 cleavages 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). 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, an 80-fold increased risk, and this risk appears to be 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. The mutation is believed to have arisen in a single ancestor approximately 21,000 to 34,000 years ago. 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. The prevalence of this abnormality in the white population is estimated to range from 0.7% to 4%.

In a group of 39 women with recurrent miscarriages, the factor V Leiden mutation was identified in nearly half, and 5 women treated with anticoagulation had normal term deliveries. Others have found an increased incidence of the factor V Leiden mutation in women with second-trimester losses, but either not at all or only a slight increase in women with first-trimester losses. In a survey of 500 consecutive women with unexplained recurrent pregnancy losses, von Willebrand’s disease, fibrinogen deficiency, deficiencies in antithrombin, protein C, and proteins S, and resistance to activated protein C were not more frequent. In a prospective cohort study, inherited deficiencies were associated with increased risks of late loss, but not for early miscarriage. Elevated homocysteine levels are a risk factor for venous thrombosis, and this condition has also been suggested to be a factor in recurrent early pregnant losses. Overall, these results are less than clear-cut.

In view of the new recognition for the clinical importance of inherited disorders that predispose to venous thrombosis, we believe that women with recurrent miscarriages who have no obvious identified cause should consider hemorologic screening. This is definitely necessary in women with a previous episode or a close family history (parent or sibling) of thromboembolism. 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. A positive result warrants treatment with low molecular weight heparin after pregnancy is diagnosed.

<table>
<thead>
<tr>
<th>Hypercoaguable Conditions</th>
<th>Thrombophilia Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antithrombin III deficiency</td>
<td>Antithrombin III</td>
</tr>
<tr>
<td>Protein C deficiency</td>
<td>Protein C</td>
</tr>
<tr>
<td>Protein S deficiency</td>
<td>Protein S</td>
</tr>
<tr>
<td>Factor V Leiden mutation</td>
<td>Activated protein C resistance ratio</td>
</tr>
<tr>
<td>Prothrombin gene mutation</td>
<td>Activated partial thromboplastin time</td>
</tr>
<tr>
<td>Antiphospholipid syndrome</td>
<td>Hexagonal activated partial thromboplastin time</td>
</tr>
</tbody>
</table>

Immunologic Problems

Autoimmunity (Self Antigens)

In autoimmunity, a humoral or cellular response is directed against a specific component of the host. The lupus anticoagulant and antiphospholipid antibodies are antiphospholipid antibodies, which arise as the result of an autoimmune disease. The lupus anticoagulant is present in a variety of clinical conditions, not just with lupus erythematosus. The antiphospholipid antibodies are directed against platelets and the vascular endothelium and cause thrombosis, spontaneous miscarriage, and fetal wastage. These antibodies block prostacyclin formation, which results in unbalanced thromboxane activity, leading to vasoconstriction and thrombosis. In several series, 10–16% of women with recurrent miscarriages have had antiphospholipid antibodies. These antibodies are also associated with fetal growth retardation and fetal death in addition to recurrent miscarriages, and when present, there is a high rate of second-trimester fetal deaths. The mechanism of pregnancy loss is probably decidual and placental insufficiency due to the thrombotic tendency. Annexin-V is a phospholipid-binding protein that inhibits coagulation; the levels of annexin-V on trophoblasts and endometrial cells are reduced in the presence of antiphospholipid antibodies.

Despite activating thrombosis, the antiphospholipid antibodies prolong the prothrombin time and the partial thromboplastin time. The activated partial thromboplastin time is a relatively sensitive screening test, but we also obtain a kaolin clotting time. The antiphospholipid antibody and lupus anticoagulant can be identified and titred by specific immunoassays; other individual antiphospholipid antibodies have not been associated with recurrent miscarriages. The antiphospholipid antibodies all produce the same clinical impact and have identical effects on clotting tests. Although the prevalence is uncertain, patients with recurrent miscarriages should be screened with the activated partial thromboplastin time, a kaolin clotting time, lupus anticoagulant, and the antiphospholipid antibody. It is not clinically helpful to measure antinuclear antibodies, IgA antibodies, or leukocyte antibodies.

Our preferred treatment for significant titers of antiphospholipid antibodies consists of the combination of low-dose aspirin (80 mg daily) and low-dose heparin as soon as...
as pregnancy is diagnosed. Studies of low-dose aspirin alone do not demonstrate a favorable effect on pregnancy outcome in either women with unexplained recurrent losses or in women with delectable anticardiolipin antibodies. Others advocate the addition of a glucocorticoid in a dose sufficient to restore the clotting studies to normal. However, the addition of glucocorticoids is not very effective in eliminating the anticardiolipin antibody, and in a randomized trial, a combination of prednisone and aspirin (100 mg daily) was no better than placebo treatment. Many of these patients develop preeclampsia, often very severe, but approximately 75% of patients with antiphospholipid antibodies will deliver a viable infant in a treated pregnancy. Because of the risks associated with anticoagulation, aspirin and heparin treatment should be confined to women with recurrent pregnancy losses who have antiphospholipid antibodies. Indeed, it can be argued that treatment should be offered only to women with antiphospholipid antibodies who have experienced both first-trimester and second-trimester losses.

Alloimmunity (Foreign Antigens)

Alloimmunity refers to all causes of pregnancy losses related to an abnormal maternal immune response to antigens on placental or fetal tissues. Normally, maintenance of pregnancy may require the formation of blocking factors (probably complexes of antibody and antigen) that prevent maternal rejection of fetal antigens. It has been argued that couples with repetitive miscarriages have an increased sharing of human leukocyte antigens (HLA), a condition that would not allow the mother to make blocking antibodies.

Immunotherapy to stimulate antibody formation has been offered to produce a favorable maternal immune response in order to protect the developing embryo. Women with recurrent miscarriages have been treated with infusions of their partner’s lymphocytes. In one study, 77% of women receiving her husband’s cells gave birth compared to 37% receiving their own cells. Critics of this study have contended that success in the control group raises the question of the adequacy of matching in the selection of control patients. Others have claimed good results with transfusion of leukocyte-rich, erythrocyte-rich donor blood (3 transfusions every 4–8 weeks) or with intravenous immunoglobulin or seminal plasma vaginal suppositories. Women have been selected for immunotherapy by HLA typing, but its use is no longer supported. Many investigators have failed to confirm that sharing of HLA antigens is found to a greater degree in couples with recurrent miscarriages. This agrees with experiments in animals where sharing of HLA antigens has not been found to affect reproduction. The sharing of genetic loci may be a broader problem that includes genetic loci critical for embryonic development, and the antigens available for measurement serve only as markers for a more fundamental genetic immunologic failure in pregnancy. There is also concern that immunization of mothers may affect placenta and fetus. There is no specific immunologic test or clinical method which will predict the need for treatment.

Immunotherapy remains experimental with potential risks of adverse consequences on the immune systems of mother and child. At least 5 randomized placebo-controlled studies have failed to demonstrate a beneficial effect of immunotherapy. In a study of women with 5 or more early miscarriages, paternal leukocyte immunization appeared to have a beneficial effect; suggesting that 25% of multiple aborters (without a live birth) would benefit. In a re-analysis of the world-wide raw data, the difference in birth rate between treatment and control groups amounted to 10%, a difference so small that it may not be clinically significant. Our recommendation is that women with an extreme number of recurrent early pregnancy losses should be referred to a center operating an immunotherapy protocol.

Summary of Immunologic Causes

Women whose recurrent miscarriages are more likely to have an immunologic cause have the following characteristics:

1. Many previous spontaneous miscarriages.
2. No recent full term pregnancies.
3. Less than 35 years old.
4. Aborted conceptus with a normal karyotype.
5. Usually at least one loss after the first-trimester.

Summary of Laboratory Evaluation and Management for Repeated Early Pregnancy Losses

<table>
<thead>
<tr>
<th>Category</th>
<th>Evaluation</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Statistics</td>
<td>Numbers review, No lab tests</td>
<td>Education and support</td>
</tr>
<tr>
<td>Genetic Factors</td>
<td>Karyotypes of both parents</td>
<td>Counseling</td>
</tr>
<tr>
<td>Environmental Factors</td>
<td>No lab tests</td>
<td>Donor gametes where appropriate</td>
</tr>
<tr>
<td>Endocrine Factors</td>
<td>TSH, prolactin, luteal phase assessment</td>
<td>Counseling</td>
</tr>
<tr>
<td>Anatomic Causes</td>
<td>Vaginal ultrasound with saline instillation, confirmed by MRI</td>
<td>Surgery</td>
</tr>
<tr>
<td>Infectious Causes</td>
<td>Cultures if clinically indicated</td>
<td>Empiric doxycycline or erythromycin</td>
</tr>
<tr>
<td>Thrombophilia</td>
<td>Screen for inherited predisposition for thrombosis</td>
<td>Low molecular weight heparin anticoagulation</td>
</tr>
<tr>
<td>Immunologic</td>
<td>Activated partial thromboplastin time, kaolin clotting time, anticardiolipin antibody, lupus anticoagulant</td>
<td>Low-dose aspirin and heparin</td>
</tr>
</tbody>
</table>

Conclusion

The patient with early miscarriages usually presents as an anxious, frustrated individual on the verge of despair. Evaluation should be spaced over several visits, allowing the clinician to establish communication and rapport with the patient. Frequent communication between the clinician and the patient during the first-trimester of the next pregnancy is essential. The emotional support that the clinician can bring to this interaction will be most useful and in some cases may be therapeutic. It should be emphasized that continued attempts at conception are rewarded with success in the majority of women (70–75%) labeled as recurrent aborters and who have no identifiable cause. Except in the case of a second-trimester loss which is associated with a poor prognosis in the subsequent pregnancy with increased risks for preterm delivery, stillbirth, and neonatal death, approximately 50% of women with histories of recurrent pregnancy losses have no identifiable abnormalities and do well in their next pregnancies. All subsequent pregnancies should be closely monitored because there is a higher rate of ectopic pregnancies in women with recurrent miscarriages.

The expectant management of spontaneous early miscarriages is now known to provide an outcome (both medical and psychologic morbidity) equivalent to surgical evacuation of the uterus, with no adverse effect on future fertility. For appropriate patients, this approach with effective support can be easier and less distressing.

Finally, the outcome of the assisted reproductive technologies is not certain in women with a large number of recurrent miscarriages. In one series of 12 women, 8 women delivered at term after in vitro fertilization. In addition, in vitro fertilization outcome is not reduced in women with positive antiphospholipid antibodies. However, in another in vitro fertilization experience, although pregnancy rates were good, a high rate of spontaneous miscarriage (50%) occurred in women with 3 or more consecutive losses. Good results have been reported with oocyte donation in women with recurrent losses. These are individual decisions, heavily
Endometriosis is a disease involving ectopic (outside the uterus) endometrial glands and stroma. In its clinical manifestations, it is a progressive disease that is a vexing problem for both patient and clinician. However, clinical studies over the past 2 decades have provided information for a better understanding of the disease and better decision-making regarding management options. This chapter will review the more recent information regarding treatment as well as what is known concerning the etiology and pathogenesis of endometriosis.

Etiology of Endometriosis

Endometriosis was described in the medical literature in the 1800s, but it was not until the 20th century that its common occurrence was appreciated. Based on clinical observation and examination of histopathologic specimens, John Sampson of Albany, New York, in 1921, suggested that peritoneal endometriosis in the pelvis arose from seedings from ovarian endometriosis. Subsequently, in 1927, he published his classic paper, “Peritoneal Endometriosis Due to Menstrual Dissemination of Endometrial Tissue Into the Peritoneal Cavity,” which introduced the term “endometriosis” and established retrograde flow of endometrial tissue through the fallopian tubes and into the abdominal cavity as the probable cause of the disease. The conclusions of Sampson have been validated by the following observations:

1. During laparoscopy, flow of blood from the fimbriated end of the tube has been observed in virtually all menstruating women.
2. Endometriosis is most commonly found in dependent portions of the pelvis, most frequently on the ovaries, the anterior and posterior cul-de-sac, and the uterosacral ligaments, followed by the posterior uterus and posterior broad ligaments.
3. Endometrial fragments from the menstrual flow can grow both in tissue culture and following injection beneath the abdominal skin, and can be retrieved from the peritoneal fluid of most menstruating women.
4. Endometriosis developed when the cervices of monkeys were transposed so that menstruation occurred into the peritoneal cavity.
5. A higher incidence of endometriosis is observed in women who have obstructions to the outward flow of the menstrual effluvium.
6. The risk of endometriosis is increased in women with shorter menstrual cycles and longer flows, characteristics that give greater opportunity for ectopic endometrial implantation.
7. Retrograde menstruation has been observed to occur more frequently in baboons with spontaneous endometriosis.

However, there may be other pathways for the development of endometriosis. Endometriosis at sites distant from the pelvis may be due to vascular or lymphatic transport of endometrial fragments. Even the common occurrence of endometriosis on the ovaries can be explained by lymphatic flow from the uterus to the ovary.

Endometriosis may occur in almost every organ of the body. For example, pulmonary endometriosis occurs and can be manifested by asymptomatic nodules or as pneumothorax, hemoptothis, or hemoptysis during menses. Urologic endometriosis is of importance because of the possibility of ureteral obstruction.

Extrapelvic endometriosis is occasionally encountered years after surgical removal of the uterus and the ovaries. This endometriosis can be hormone-resistant, and we can only speculate as to why it grows. One possibility is transplantation of endometrial implants during the original surgery, or activation of residual disease. Another possibility is transformation by metaplasia of other tissue or activation of embryonic rest tissue. When the endometriosis is hormone-sensitive, a good possibility is that an ovarian remnant was left behind during the complicated surgery, and this allows continuing hormonal stimulation of residual endometriosis.

There are case reports of endometriosis in men who received treatment with estrogen, and therefore, another possible cause of endometriosis is the transformation of coelomic epithelium into endometrioid-type glands as a result of unspecified stimuli. The following arguments can be used to support the coelomic metaplasia theory:

1. Endometriosis occurs in adolescent girls in the absence of müllerian anomalies, and it can be discovered a few years after menarche before many menstrual cycles have been experienced.
2. Endometriosis has been reported in a prepubertal girl.
3. Endometriosis has been encountered in women who never menstruated.
4. Endometriosis in unusual sites, such as thumb, thigh, or knee, can be explained by the fact that mesenchymal limb buds develop adjacent to coelomic epithelium during early embryogenesis.
5. Endometriosis does occur in men, usually associated with high-dose estrogen treatment.

Because many women have reflux seeding of menstrual debris into the peritoneal cavity, and not all develop endometriosis, there may be genetic or immunologic factors that influence the susceptibility of a woman to the disease. Simpson and coworkers reported 6.9% of first-degree relatives of patients with endometriosis had the disease, compared with 1.0% in a control group.

Monozygotic twins are markedly concordant for endometriosis. A worldwide collaborative project (The Oxford Endometriosis Gene Study) has been organized to identify a genetic basis for endometriosis. A web site is available and can be used to provide access to the study for affected families: http://www.medicine.ox.ac.uk/ndog/oxegene/oxegene.htm.

Dmoski and coworkers demonstrated that monkeys with endometriosis had decreased cellular immunity to endometrial tissue, suggesting that specific immunologic defects can render some individuals susceptible to endometriosis. Others have found an increased prevalence of humoral antibodies directed against endometrial and ovarian tissue in the sera of women with endometriosis. In addition, women with endometriosis demonstrate a decrease in various measurements of immune response.

A consideration of the etiologic theories regarding endometriosis leads to the conclusion that all of these mechanisms can contribute to the clinical problem in an individual patient, and the degree of contribution for each probably varies from patient to patient. Endometrial cells can be spread by mechanical means, or perhaps...
Prevalence of Endometriosis

Widely varying figures for the prevalence of endometriosis have been published, and a rough estimate is that 3–10% of women in the reproductive age group and 25–35% of infertile women have endometriosis. About 4 per 1000 women age 15–64 are hospitalized with endometriosis each year, slightly more than those admitted with breast cancer. The common perceptions that endometriosis only occurs in goal-oriented women over the age of 30 and is not found often in black women have now been discredited. Whereas endometriosis essentially does not occur before menarche, there are increasing reports of its occurrence in the teenage years. A number of these cases involve anatomic abnormalities that obstruct the outflow tract. Endometriosis is not confined to nulliparous women, and physicians should be alert to the presence of endometriosis in cases of secondary infertility.

Diagnosis of Endometriosis

Endometriosis should be suspected in any woman complaining of infertility. Suspicion is heightened when there are also complaints of dysmenorrhea and dyspareunia.

Symptoms and Signs

Dysmenorrhea and deep dyspareunia are even more suggestive of endometriosis if the symptoms begin after years of relatively pain-free menses and coitus. It should be recognized, however, that many women who have endometriosis are asymptomatic. A common observation is that some women with extensive endometriosis have little or no pain, whereas others with only minimal endometriosis complain of severe pain. Very severe pain, however, is associated with deeply infiltrating endometriosis. Thus, the lack of correlation between pain and visible endometriosis is because the degree of pain is determined by the depth of infiltration. In addition, it has been our observation that midline disease is usually more symptomatic than lesions that are laterally placed.

Pain can be diffuse in the pelvis or it can be more localized, often in the area of the rectum. Symptoms also can arise from rectal, ureteral, or bladder involvement with endometriosis, and can be present throughout the month. Blockage of the ureter can occur, and urinary tract symptoms should be investigated with urologic and radiologic techniques. Low back pain, too, may be due to endometriosis. An association of endometriosis and premenstrual spotting has been suggested, but in most cases menstrual dysfunction is not increased with endometriosis. An association between galactorrhea and endometriosis has been claimed, but baseline elevations of prolactin are not higher in patients with endometriosis than in normal women.

The CA-125 Assay

CA-125 is a cell surface antigen found on derivatives of the coelomic epithelium (which includes endometrium), and it is a useful marker in the monitoring of women with epithelial ovarian carcinoma. In addition, serum CA-125 levels are often elevated in patients with endometriosis and correlate with both the degree of disease and the response to treatment. The sensitivity of this assay is too low to use it as a screening test, but it can be a marker of response to treatment and for recurrence; however, elevated levels which suppress during medical treatment often promptly return to pretreatment concentrations immediately after cessation of therapy, limiting its clinical usefulness. Serum CA-125 determinations may be able to differentiate endometriotic from nonendometriotic benign adenexal cysts, especially when combined with transvaginal ultrasonography. Note that CA-125 levels can be elevated by early pregnancy, acute pelvic inflammatory disease, leiomyomata, and menstruation.

Examination

We recommend that pelvic examination be performed at the time of menses when tenderness is easier to detect. The uterus is often in fixed retroversion and the ovaries may be enlarged. However, retroversion of the uterus is not an etiologic factor, and prophyllactic uterine suspension is no longer recommended. Nodularity (which is usually tender) of the uterosacral ligaments and cul-de-sac can be found in one-third of patients with endometriosis. The diagnosis almost always should be confirmed by laparoscopy before treatment is initiated. Minimal findings, such as slight beading and tenderness of the uterosacral ligaments in the young, asymptomatic patient, can be treated initially, however, with combined, low dose oral contraceptives.

The appearance of endometriosis is quite varied. All too often the clinician fails to observe endometrial lesions because of a preconceived expectation limited to the classic blue or black powder burn appearance. Lesions can be red, black, blue, or white and nonpigmented. Biopsies from visibly normal peritoneum can contain endometriosis in 6–13% of infertile women; however, the clinical significance of this presence is uncertain (and, in our view, unlikely to be important). Adhesions, peritoneal defects, and tan, creamy, fresh-appearing endometrium also can be observed. The dark pigmented lesions are later consequences of tissue bleeding responses to cyclic hormones. The ovary is the most common site for both implants and adhesions, followed by widespread distribution, anteriorly and posteriorly, over the broad ligament and cul-de-sac.

A Classification System

Because both treatment and prognosis are determined to some extent by the severity of the disease, it is desirable to have a uniform system of classification that takes into account both the extent and severity of the disease. A uniform classification is also crucial for comparing the results of different treatments. The American Society for Reproductive Medicine developed a classification system based on findings at laparoscopy or laparotomy. However, there were weaknesses in the classification system, especially the fact that it was based upon the arbitrary impressions of the clinician. A revised classification system was produced to standardize the documentation of findings in patients who have pelvic pain and endometriosis, and forms are available from the Society. There can be a high intraobserver and interobserver variability in the evaluation of endometriosis using any classification system. Therefore, efforts must continue to provide an accurate method for staging.

Endometriosis and Infertility

When endometriosis involves the ovaries and causes adhesions that block tubal motility and pickup of the egg, there is no question of its role in causing mechanical interference with fertility. Less secure is the information on the role of peritoneal endometriosis on fertility. Many physicians believe that even minimal endometriosis can cause infertility, impairing tubal function or gamete quality. This argument has been weakened by a failure to find benefit from medical treatment of infertility associated with minimal to mild endometriosis.

The absence of benefit from therapy, however, could represent a problem with the treatment rather than a lack of association between infertility and endometriosis. Endometriosis diagnosed by laparoscopy is reported in a higher proportion of infertile women (38.5%) compared with fertile women (5.2%). Moreover, fecundity rates in women with endometriosis tend to be lower than the normal fecundity rate. However, the long-term cumulative pregnancy rates are very high in women who have minimal to mild endometriosis and are not treated.

If minimal or mild endometriosis does affect fertility, what are the mechanisms? Certainly, dyspareunia secondary to endometriosis could play a role.

Another mediator could be prostaglandins produced by the implants, which could, in turn, affect tubal motility, or folliculogenesis and corpus luteum function. Patients with endometriosis have been reported to have an increase in both the volume of peritoneal fluid and the concentration of thromboxane B2 and 6-keto-prostaglandin F1α in the fluid. Others, however, found neither an increase in peritoneal fluid nor an increase in concentration of peritoneal fluid prostaglandin E2, prostaglandin...
Implants of endometriosis react to steroid hormones in a manner somewhat, but not exactly, continuously for 6–12 months. Estrogen (conjugated estrogens 1.25 mg or estradiol 2.0 mg daily for 1 week) is added if breakthrough bleeding occurs. The treatment matter little which low dose monophasic product is used to accomplish the conversion of endometrial implants into decidualized cells associated with a few inactive

Until the late 1970s the most important alternative to conservative surgery was the use of combination oral contraceptives taken in a continuous fashion. Medical Treatment of Endometriosis

Phagocytosis of sperm by macrophages could be one mechanism of action. However, patients with and without endometriosis have the same number of motile sperm recoverable from the peritoneal cavity. Peritoneal macrophages from women with endometriosis secrete interleukin-1 which is toxic to mouse embryos. In addition, cytokines are elevated in the peritoneal fluid, which could recruit macrophages and lymphocytes and perpetuate the inflammatory reaction. Macrophages and endometrial cells also produce vascular endothelial growth factor (VEGF), an important angiogenic factor involved in endometrial proliferation.

Endometriosis has been implicated as a factor in disordered follicle growth, ovulatory dysfunction, and failure of embryo development. Ultrasoundography studies suggest that there is some retardation in growth of follicles in women with endometriosis. Luteinized unruptured follicle syndrome (LUF) in which the oocyte is not released at the time of follicle rupture (or there is failure of the follicle to rupture) has been suggested as a cause of both unexplained infertility and infertility secondary to endometriosis. Although an unruptured follicle can occur in women, there currently is no impressive evidence that this syndrome is secondary to endometriosis or is even a cause of infertility.

The question of how minimal or mild endometriosis can affect fertility now has been superseded by the question of whether there is any effect of mild endometriosis on fertility. More importantly, should endometriosis be treated if the complaint is infertility and not pain? Many articles purporting to show that therapy overcomes endometriosis-associated infertility are flawed by lack of control groups and the failure to use life-table analyses. Moreover, expectant management of mild endometriosis is rewarded with reasonable pregnancy rates that are comparable to those obtained with treatment. A cumulative pregnancy rate after 5 years of 90% has been reported in women not treated for minimal or mild endometriosis.

Whereas the studies cited above strongly suggest that medical or surgical treatment of mild endometriosis may not be worthwhile, there are those who champion fulguration treatment under laparoscopic visualization. In a randomized trial involving 341 infertile women with minimal or mild endometriosis, comparing diagnostic laparoscopy with laparoscopic resection or ablation of visible endometriosis, the surgically-treated group experienced a 48% pregnancy rate (over 9 months) compared with 35% in the non-treated group. Of note was the observation that this improvement only occurred in women with blue or black lesions, not with red lesions. Furthermore, an argument can be made that active treatment of even mild endometriosis is warranted because endometriosis is often a progressive disease.

The increase (only a modest one) in the pregnancy rate associated with laparoscopic treatment is consistent with the good pregnancy rates associated with expectant management of mild endometriosis. Nevertheless, it is currently recommended that resection or ablation should be immediately performed when minimal or mild endometriosis is diagnosed during laparoscopy for infertility. Moderate and severe endometriosis should be treated surgically at the time of diagnosis. If pregnancy is not achieved after surgery, infertility should be treated first with superovulation and IUI, and then with IVF.

Surgical Treatment of Endometriosis

In contrast to the dispute over the proper treatment of mild endometriosis, there is little doubt that adhesive disease associated with endometriosis, or large (>2 cm) endometriomas, is best treated by surgery. That subject of surgery should be to restore normal anatomical relationships and to excise or fulgurate as much of the endometriosis as possible. Removal of severely diseased adnexa when the other side is more normal produces better results than attempts to do major repairs. Presacral neuroectomy does not enhance fertility, although many surgeons advocate it to alleviate dysmenorrhea. This may be less compelling now that progesterin inhibitors are available to accomplish the same purpose. A careful study of presacral neuroectomy concluded that this procedure is only indicated in patients with pain limited to the midline area.

The success of surgery in relieving infertility is directly related to the severity of endometriosis. Patients with moderate disease can expect a pregnancy success of approximately 60%, whereas the comparable figure is 35% in those with severe disease. As noted above, surgical treatment of blue or black lesions yielded a slightly higher (13%) pregnancy rate compared to expectant management in a randomized trial. Of note was the observation that this improvement only occurred in women with blue or black lesions, not red lesions. There is support for selective use of medical treatment for 2–3 months following laparoscopy and prior to conservative surgery, especially in patients with pain due to major disease. Preoperative treatment aids surgery by softening endometrial implants.

Postoperative use of hormones has been the subject of greater controversy. The highest pregnancy rates following conservative surgery occur in the first year after surgery, and most physicians have been reluctant to use hormones that prevent pregnancy even for a few months. If pregnancy does not occur within 2 years of surgery for endometriosis, the chances are poor that pregnancy will occur. The recurrence rate reported for endometriosis after surgery is 20% within 5 years, and when it does recur, second surgeries to aid fertility have only a limited chance for success. When the objective of laparoscopic surgery is symptomatic relief and not pregnancy, a return of symptoms is delayed with as little as 6 months of postoperative medical treatment.

The type of surgery that we have been discussing is labeled "conservative" to indicate that reproductive function is maintained. When endometriomas are removed a vigorous attempt should be made to leave behind any normal ovarian tissue. Even one-tenth of an ovary can be enough to preserve function and fertility. Conservative surgery can be accomplished by laparoscopy which decreases costs and morbidity, yet provides results that are as efficacious in all stages of disease as laparotomy.

"Conservative" surgery is in contradistinction to "radical" surgery, which includes hysterectomy and usually bilateral salpingo-oophorectomy. When radical surgery is performed, an unovulated ovary can be preserved in some cases if all of the nonovarian endometriosis is removed by fulguration or excision. However, this does provide a risk for recurrent disease. Patients who undergo hysterectomy with ovarian conservation have a 6-fold increased risk of developing recurrent symptoms compared with women who have oophorectomy.

Medical Treatment of Endometriosis

Although hormonal therapy of infertility associated with endometriosis is not of proven value, medical therapy for dysmenorrhea, dyspareunia, and pelvic pain associated with endometriosis is very successful (although relief may be short-term). The various agents used are comparable in terms of efficacy, and choice of treatment is determined by cost and side effects. Implants of endometriosis react to steroid hormones in a manner somewhat, but not exactly, similar to normally stimulated endometrium. However, endometriotic tissue displays histologic differences and biochemical differences, including enzyme activity and receptor levels that differ in concentration and response compared to normal endometrium. Nevertheless, estrogen stimulates growth of the implants. For this reason, endometriosis usually regresses following menopause and is usually not found prior to menarche unless there is a blockage of the outflow tract.

Hormonal therapy is designed to interrupt the cycle of stimulation and bleeding. An early approach was the use of massive doses of diethylstilbestrol (DES), which, because of variable success, the risk of affecting the fetus, and side effects of severe bleeding and nausea, is now of only historical interest. Treatment with androgens (methyltestosterone linguits 5–10 mg/day) can provide only transient relief of the pain of endometriosis, and its effect on infertility appears to be negligible. In addition, ovulation can occur while on treatment, and there is a risk of exposure of the fetus to the androgen.
with oral contraceptives was called pseudopregnancy because of the amenorrhea and the decidualization of the endometrial tissue induced by the estrogen-progestin combination. It also reflected the commonly held belief that pregnancy can improve endometriosis, a belief that has been disputed. The side effects of treatment are those associated with oral contraceptives (Chapter 22). Pregnancy rates after stopping medication are reported to be in the 40–50% range. Whereas published recurrence rates are not excessive, this therapy, as with all medical treatment for endometriosis, must be viewed as suppressive rather than curative.

Treatment With Danazol

The golden age for danazol has passed. Although its expense and side effects seemed a reasonable tradeoff for an effective treatment, it is now apparent that danazol is no more effective than the other medications used to treat endometriosis. In distinction to the pseudopregnancy induced by oral contraceptives, danazol produces what has been incorrectly termed as pseudomenopause. Danazol is an oxazolone derivative of the synthetic steroid 17a-ethinylestradiol. It originally was thought to exert its effect solely by inhibition of pituitary gonadotropins. Although danazol can decrease follicle-stimulating hormone (FSH) and luteinizing hormone (LH) in castrated individuals, it does not alter basal gonadotropin concentrations in premenopausal women. It does, however, eliminate the midcycle surge of FSH and LH. Danazol inhibits steroidogenesis in the human corpus luteum. Danazol is metabolized to at least 60 different products, some of which may contribute to its many effects. The multiple effects of danazol produce a high-androgen, low-estrogen environment that does not support the growth of endometriosis, and the amenorrhea that is produced prevents new seeding from the uterus into the peritoneal cavity.

The side effects of danazol are related both to the hypoestrogenic environment it creates and to its androgenic properties. The most common side effects are weight gain, fluid retention, fatigue, decreased breast size, acne, oily skin, growth of facial hair, atrophic vaginitis, hot flushes, muscle cramps, and emotional lability. Some of these side effects occur in approximately 80% of women who are taking danazol, but less than 10% find the side effects sufficiently troublesome to warrant discontinuation of the drug. Because danazol has been associated with the development of uterine and vulval pseudohermaphroditism, it should not be given if there is the possibility of pregnancy. The androgenic action of danazol can irreversibly deepen the voice. It is worth inquiring whether singing is an important part of your patient’s life before prescribing this drug.

Danazol is metabolized largely in the liver, and in some patients it causes hepatocellular damage. Its use, therefore, is contraindicated in women with liver disease. Furthermore, liver enzymes should be monitored during treatment with danazol. The fluid retention that is often associated with danazol makes it dangerous to use when there is severe hypertension, congestive heart failure, or impaired renal function. It can produce increased cholesterol and low-density lipoprotein levels and decreased levels of high-density lipoprotein. It is unlikely that these short-term effects on lipids and lipoproteins are clinically important. The drug has been used to treat autoimmune disease, but it is not known if this action plays a role in its effects on endometriosis.

It should be noted again that medical treatment of mild endometriosis has been called into question because women with untreated mild endometriosis have pregnancy rates equal to women who have received medical treatment for the endometriosis. Thus, danazol treatment of infertility associated with mild endometriosis has no scientific support.

On the other hand, danazol is useful to relieve the pain of endometriosis and to prevent progression of the disease. Pain relief is obtained in 90% of patients. The usual dose is two 200 mg tablets twice a day (although some claim that spacing the drug at 6-hour intervals may be more effective) for 6 months. Dnoewski and Cohen reviewed 99 women who completed danazol treatment for a period of 3–18 months (average 6 months) and who were reevaluated an average of 37 months later. During the course of treatment all the patients had asymptomatic improvement, and the majority (85%) were clinically improved. At the time of the reevaluation, however, approximately one-third were symptomatic and had clinical findings suggestive of recurrent endometriosis. In the majority of these patients, the symptoms recurred within the first year after discontinuation of the drug.

The success of danazol treatment is greatest in cases of peritoneal endometriosis or those with small lesions of the ovary. Endometriomas larger than 1.0 cm are less likely to respond to danazol, although quite surprising regression of endometriomas larger than 1.0 cm is sometimes seen. Because of the significant side effects encountered with danazol, and its cost, there has been a trend toward the use of lower doses than the usual 800 mg daily. However, doses below 600 mg may be less effective. The occurrence of amenorrhea appears to be correlated with improved outcome, and this is more consistently obtained at the 800 mg level. Others, however, do not believe that amenorrhea is an important consideration because many of the patients will bleed from an atrophic endometrium.

There is a general perception, but only limited experimental evidence, that danazol is more effective than oral contraceptives for the treatment of endometriosis. Noble and Letchworth compared danazol with a high dose estrogen-progestin combination (Enovid). The dose of both danazol and the oral contraceptive was increased until the patients became amenorrheic. One of 25 patients taking danazol could not complete 5 months of treatment, whereas 7 of 17 (41%) of the group taking oral contraceptives dropped out because of side effects. Danazol was more effective than estrogen-progestin in relieving symptoms, and laparoscopic assessment showed much better results with danazol. Seven of 12 danazol-treated patients became pregnant compared with 4 of the 10 women who had taken oral contraceptives. Unfortunately, the study size was too small to allow statistical analysis.

Treatment With a Progestational Agent

Both oral and injectable medroxyprogesterone acetate have been effective in treating endometriosis by causing decidualization and subsequent atrophy of endometrial tissue. Medroxyprogesterone acetate in an oral dose of 30 mg daily has been demonstrated to be as effective as danazol in treating endometriosis. Similar results have been obtained with higher doses. Although not studied, it is likely that lower doses (10–20 mg daily) would be effective. For this reason and because it is more cost-effective and there are fewer side effects, medroxyprogesterone acetate is often the first choice for medical treatment of endometriosis. High doses of medroxyprogesterone can adversely affect the lipoprotein profile; there is no reason to use a dose greater than 30 mg/day. Megestrol acetate has been shown to increase until the patients become amenorrheic. One of 25 patients taking danazol could not complete 5 months of treatment, whereas 7 of 17 (41%) of the group taking oral contraceptives dropped out because of side effects. Danazol was more effective than estrogen-progestin in relieving symptoms, and laparoscopic assessment showed much better results with danazol. Seven of 12 danazol-treated patients became pregnant compared with 4 of the 10 women who had taken oral contraceptives. Unfortunately, the study size was too small to allow statistical analysis.

Side effects include weight gain, fluid retention, and breakthrough bleeding. Breakthrough bleeding is a common occurrence although it is usually cleared by short-term (7 days) administration of estrogen. Depression is a significant problem, and both patient and physician should be alert for its development. The usefulness of depot-medroxyprogesterone acetate (150 mg im every 3 months) in infertile patients is limited by the varying length of time it takes for ovulation to resume after discontinuation of therapy. This is not a problem with oral administration.

Medroxyprogesterone acetate, like danazol, can relieve the symptoms of endometriosis, but it is not effective in treating infertility. In a prospective, randomized clinical trial, there was no difference in pregnancy rates following treatment with medroxyprogesterone acetate (100 mg/day) or placebo.

Treatment With GnRH Agonists

Gonadotropin-releasing hormone (GnRH) has a short half-life because it is rapidly cleaved between amino acids 5–6, 6–7, and 9–10. Analogues of GnRH have been produced by altering the amino acids at these positions. Substitutions of amino acids at the 6 position and/or replacement of the C-terminal glycine-amide (inhibiting
A long-acting GnRH agonist can create a pseudomenopause for the treatment of endometriosis. 12, 13 At the end of 2-4 weeks of daily administration of the agonist, estrogen levels will decrease to those found in oophorectomized women. Dosage can be adjusted by monitoring serum estradiol levels; the best therapeutic effect is associated with a range of 20–40 pg/mL. Thus, the “medical oophorectomy” caused by the continuous use of a GnRH agonist has provided a new approach to the treatment of endometriosis. Excellent, large, well-designed studies (with advanced disease in nearly 50% of the patients) have compared GnRH agonist therapy (with various agents) with danazol. 13, 14, 15, 16, 17 The results in terms of reduction of disease (as demonstrated by post-treatment laparoscopies) and pregnancy rates have been the same with either treatment. However, GnRH agonist treatment does not have an adverse impact on serum lipids and lipoproteins compared to that observed with danazol. 18 An experimental comparison of agonist and progesterin treatment in monkeys concluded that the progesterin was just as effective as the agonist. 12

GnRH Agonists in Clinical Use

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native GnRH</td>
<td>pGluHisTrpSerTyrGly</td>
<td>Leu</td>
<td>Arg</td>
<td>Pro</td>
<td>Gly-NH₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leuprolide</td>
<td>D-Leu</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buserelin</td>
<td>D-Ser (tertiary butanol)</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nafarelin</td>
<td>D-Naphthylalanine (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histrelin</td>
<td>D-His (tertiary benzyl)</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goserelin</td>
<td>D-Ser (tertiary butanol)</td>
<td>Aza-Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deslorelin</td>
<td>D-Trp</td>
<td>NH-Ethylamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptorelin</td>
<td>D-Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An uneventful pregnancy with delivery of a normal infant has been reported despite GnRH agonist treatment (monthly treatment with a long-acting agonist was continued through the first two months of pregnancy). 14

As with all other drug therapies of endometriosis, the GnRH agonist provides suppression rather than cure of the disease. The long-term consequences of the hypoestrogenic state on calcium metabolism and bone are of concern. Therefore, treatment is usually limited to 6 months to avoid bone loss, although even during this time period there can be a 6–8% decrease in trabecular bone density. This short-term bone loss is reversed after cessation of therapy, but many patients have not regained the bone that was lost, as much as one year later. 14, 15, 16 A study with longer follow-up indicated that bone density was restored by 2 years after 6 months of GnRH agonist treatment. 16 Because long-term therapy carries with it the concern over a lasting impact on the risk of osteoporosis, consideration should be given to add-back therapy. The addition of a 19-nortestosterone progestin as add-back treatment is effective in decreasing bone loss. 17, 18 A postmenopausal estrogen-progesterin program can be utilized (conjugated estrogens 0.625 mg daily and medroxyprogesterone acetate 2.5 mg daily or any of the other estrogen-progesterin combinations available). This combined add-back treatment is effective for endometriosis and prevents the hypoestrogenic symptoms (especially hot flushes and vaginal dryness) associated with GnRH agonist therapy, including loss of bone. 19 Other effective options for add-back therapy to prevent bone loss are tibolone (2.5 mg daily) and bisphosphonate treatment. 19 Endometriosis by itself is not a cause of accelerated bone loss. 19, 20, 21

In well-designed trials, treatment with GnRH agonist has not increased pregnancy rates in women with infertility associated with minimal or mild endometriosis. 22

Treatment With Gestrinone

Gestrinone, a 19-nortestosterone derivative, decreases the secretion of FSH and LH and has the advantage of requiring only twice a week administration. Used in Europe, it has been as effective as danazol in treating endometriosis. 23 The clinical side effects are also similar to those seen with danazol. As with other hormonal treatments, gestrinone has been ineffective for the infertility associated with endometriosis. 23

Recurrence of Endometriosis

Endometriosis tends to recur unless definitive surgery (total hysterectomy and bilateral oophorectomy) is performed. The recurrence rate with medical treatment is approximately 5–20% per year (reaching an overall cumulative rate at 5 years as much as 40%). The recurrence rates 5 years after women were treated with various GnRH agonists were 37% for minimal disease and 74% for severe disease. 24 After 7 years, 56% of all treated women had a recurrence. In women treated for pelvic pain, the symptoms usually return within a month after cessation of medical therapy. 25 For a period of time after medical treatment, however, the intensity of symptoms is less severe. 26 It should be emphasized that pain relief, pregnancy rates, and recurrence rates are similar with all methods of medical treatment. A randomized clinical trial comparing a low-dose oral contraceptive with a GnRH agonist measured no difference in the recurrence of pain. 27 The recurrence rates after treatment with GnRH agonist are similar to those after danazol, and both are greater than that obtained with surgical excision. After conservative laparoscopy, approximately 10% of patients experience a recurrence within 1 year, and 20% within 5 years. 28, 29

Speculation regarding the reason for recurrence focuses on endometriosis (perhaps microscopic) which escaped detection, incomplete treatment, or reestablishment of primary disease by whatever mechanism is responsible. The treatment choices and results are no different than when originally confronted. Successful treatment with total pelvic irradiation has been reported in rare patients who have repeatedly failed standard therapy. 25

Hormone Treatment After Surgery

Definitive surgery for severe endometriosis, which includes abdominal hysterectomy and bilateral salpingo-oophorectomy as well as resection of all endometriosis, is the only cure for the disease. If oophorectomy is performed, estrogen-progesterin therapy at usual doses can be started immediately postoperatively with an essentially negligible risk of inducing growth of residual endometriosis or a return of pain. 30, 31 The addition of a progestational agent is strongly recommended because of reported cases of adenocarcinoma in endometriosis tissue in women treated with...
Long-Term Medical Therapy

Long-term medical therapy without surgery is useful in patients with severe symptoms but with little in the way of palpable findings. Before undertaking prolonged therapy, diagnosis should be established by laparoscopy. Prolonged medical therapy also is indicated if symptoms recur after conservative surgery.

Prevention of Infertility

A common clinical problem is the incidental finding at surgery of mild endometriosis in a young woman who has no immediate interest in pregnancy. As previously stated, we favor immediate surgical treatment of visible endometriosis when the opportunity presents itself. Consideration should be given, however, to prophylactic therapy. Combination oral contraceptives to decrease the risk of further seeding are appropriate for treatment of very mild disease, for example a few implants in the cul-de-sac. More advanced disease should be treated with 6 months of medroxyprogesterone acetate, a GnRH agonist, or danazol, followed by oral contraceptives to decrease the risk of progression of the disease. Although not well documented, clinical experience has suggested that continuous oral contraceptives are more effective as prophylaxis than the usual cyclic regimen. Indeed, we favor the continuous regimen without a break for this purpose.

As with all medical treatments, as time passes after treatment, endometriosis emerges. Although spontaneous regression occurs, endometriosis is usually progressive. This is reduced in women currently using oral contraceptives; however, effective prophylaxis requires long-term treatment. Nevertheless, treatment of endometriosis does have a beneficial impact on the natural history of the disease. Consideration should be given to the prophylactic use of oral contraceptives in women with impressive family histories of endometriosis.

Endometriosis and Spontaneous Miscarriage

Endometriosis has been purported to be associated with an increased risk of spontaneous miscarriage, a risk that is said to be substantially lessened by either medical or surgical treatment. In appropriately controlled studies, however, the miscarriage rate was in the normal range in women with endometriosis who were not treated, and it is likely that previous studies were flawed by their choice of control miscarriage rates.

Endometriosis and Ovulation

The frequency of anovulation and luteal phase defects is similar in women with and without endometriosis. Thus, a woman need not be ovulating in order to have endometriosis. One report has suggested, however, that the success of ovulation induction in women with endometriosis is enhanced by prior treatment with danazol.

Endometriosis and Assisted Reproduction

The use of superovulation with intrauterine insemination (Chapter 30 and Chapter 31) has been reported to increase fecundity rates (as high as 5- or 6-fold) in women with infertility associated with endometriosis. However, although superovulation raised fecundity rates, it did not raise cumulative pregnancy rates. This treatment may accelerate the occurrence of pregnancy without changing overall fertility.

A Patient Support Organization

The Endometriosis Association is an international organization that provides education and support for women with endometriosis. Their information prepared for teenagers is especially helpful.

Chapter References


146. Tummoos IS, Asher LJ, Martin JBB, Tuluani T, Randomized controlled trial of suprovulation and insemination for infertility associated with minimal or mild endometriosis, Fertil Steril 68:8, 1997.


The perception of the degree of male involvement in infertility has undergone a number of revisions during the past 50 years. Initially, infertility was considered primarily a female problem. This notion gave way to the realization that 40–50% of infertility is wholly or in part due to a male factor. More recently, there has been a trend to redefine, in a downward direction, the lower limit of “normal” for a sperm count. Thus, some men who in the past would have been categorized as subfertile now are considered normal.

Despite these changes, there is no doubt that a substantial percentage of infertility is due to deficiencies in the semen. For this reason it is important to be knowledgeable concerning male infertility. It usually depends on the woman's clinician to order and review the results of the initial semen analysis, and to determine the need for urologic consultation. This chapter will consider the analysis of semen and other tests of sperm function, indicate factors responsible for abnormalities of the semen, and consider available treatments for problems of male infertility, including artificial insemination. Influencing the entire field of male infertility is the striking success achieved with intracytoplasmic sperm injection (ICSI), the direct injection of a single spermatozoon or spermatid into the cytoplasm of an oocyte. Men who in the past have been considered hopelessly infertile now have reasonable chances for fertility with the use of ICSI.

Regulation of the Testes

The testes have 2 distinct components, the seminiferous tubules (site of spermatogenesis) and the Leydig cells (source of testosterone). The function of these 2 components requires both pituitary gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). The primary effect of LH is to stimulate the synthesis and secretion of testosterone by Leydig cells (about 5–10 mg per day), an effect that is enhanced by FSH, which also binds to Leydig cells and increases the number of LH receptors on the cells. Increasing levels of testosterone, in turn, inhibit LH secretion, acutely through the hypothalamus and chronically at the pituitary level. This negative feedback action does not require aromatization to estrogen. In men virtually all the estrone and estradiol present is derived from androstenedione and testosterone; there is essentially no direct secretion of estrogen.

FSH, in conjunction with testosterone, acts on the seminiferous tubules to stimulate spermatogenesis. This effect may be mediated by activation of Sertoli cell function, which is controlled by both FSH and testosterone. FSH binds to Sertoli cells and stimulates the production of several proteins, chief of which is ABP, the androgen-binding protein. Spermatogenesis requires a very high local concentration of testosterone and dihydrotestosterone, 50 times higher than that present in the circulation and greater than can be administered exogenously. The ABP is secreted into the tubule lumen and binds testosterone and dihydrotestosterone as they diffuse into the lumen, concentrating the androgens in the seminiferous epithelium for spermatogenesis and in the epididymis for sperm maturation. Spermatogenesis and fertility can occur in men who have an inactivating mutation in the FSH receptor gene, indicating that testosterone can compensate for a deficiency in FSH.

In contrast to the effects of testosterone on LH, steroid hormones at physiologic levels do not suppress FSH secretion. Orchectomy is followed, however, by a rise in FSH levels. This phenomenon led to the discovery of inhibin. Inhibin B is synthesized in the Sertoli cells in response to FSH, secreted into the circulation, and specifically inhibits FSH secretion in the pituitary. Inhibin has been found in seminal fluid, spermatozoa, Sertoli cells, and Leydig cells. Inhibin A is absent in the circulation of males. The secretion of inhibin B is further modulated by LH, human chorionic gonadotropin (HCG), and testosterone. Undoubtedly, autocrine/paracrine regulation by growth factors and local peptides is involved in a system similar to the complex interaction in the ovarian follicle. 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.
Semen Analysis

An abstinence period of 2–3 days prior to semen collection is adequate, although some urologists favor 5 days. Increasing the ejaculatory frequency reduces the volume and count but has no significant impact on quality (morphology and motility). The specimen should be collected directly into a clean container and not into a condom because the latter contains spermicidal agents. Sheathes are available that do not contain a spermicidal agent, and they can be used during intercourse if the man cannot, or will not, obtain a specimen by masturbation. Collection of a specimen by withdrawal runs the risk of losing the first part of the specimen that contains the highest concentration of sperm. Ideally, the specimen should be collected in a private room near the laboratory. If collected at home, the specimen should be protected from the cold and delivered to the laboratory within 1 hour of collection.

Semen liquefaction, which occurs 20–30 minutes after ejaculation, is a necessary prerequisite for doing an accurate analysis. On occasion, a specimen does not undergo normal liquefaction or is abnormally viscid, and, if this is associated with a poor postcoital test, it may be a factor in infertility. Techniques used to break up a viscid specimen in preparation for doing a sperm count or for artificial insemination include mechanically dispersing the gel by running the semen repeatedly through a number 19 needle, collecting the semen as a split ejaculate because the first part may be less gelatinous, or, rarely, treating the semen with a proteolytic enzyme. If the specimen does not undergo normal liquefaction or is abnormally viscid, and, if this is associated with a poor postcoital test, it may be a factor in infertility. Techniques used to break up a viscid specimen in preparation for doing a sperm count or for artificial insemination include mechanically dispersing the gel by running the semen repeatedly through a number 19 needle, collecting the semen as a split ejaculate because the first part may be less gelatinous, or, rarely, treating the semen with a proteolytic enzyme.

It has been suggested that the average sperm count has been decreasing over the past 50 years. Environmental toxins, and specifically chemicals with estrogen-like activity (xenoestrogens), have been blamed for this change. However, many studies have failed to document a decline. Most importantly, infertility has not increased significantly during the periods studied, and this argues that even if a decrease in sperm count has occurred, it is not clinically meaningful.

The World Health Organization (WHO) suggests the following for normal values, but these should be viewed as rough guidelines only.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>2.0 mL or more</td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>20 million/mL or more</td>
</tr>
<tr>
<td>Motility</td>
<td>50% or more with forward progression, or</td>
</tr>
<tr>
<td></td>
<td>25% or more with rapid progression within 60 minutes of ejaculation</td>
</tr>
<tr>
<td>Morphology</td>
<td>30% or more normal forms</td>
</tr>
<tr>
<td>White blood cells</td>
<td>fewer than 1 million/mL</td>
</tr>
<tr>
<td>Immunobead test</td>
<td>fewer than 20% spermatozoa with adherent particles</td>
</tr>
<tr>
<td>SpermMar test</td>
<td>fewer than 10% spermatozoa with adherent particles</td>
</tr>
</tbody>
</table>

Defining male infertility on the basis of sperm count or motility is more complicated than suggested by the above criteria. Whereas it is still true that very poor sperm counts (less than 5 million/mL) and very low motilities (less than 20%) indicate compromised fertility, even these low values or lower can, on occasion, be associated...
with a pregnancy. Adding to the difficulties of diagnosing male infertility is the variability in count and motility that can be seen in successive semen specimens from the same individual. This is the basis of a fundamental recommendation: at least two, and preferably three, samples should be screened before an individual can be categorized as potentially fertile, subfertile, or infertile (azoospermia). Because it can take 2.5 months for the tests to recover from an insult, it is reasonable to space the specimens over a longer period of time.

There have been conflicting statements in textbooks and articles pertaining to the lower limit of normal for the sperm count. The most commonly cited figures are 60 million/mL, 40 million/mL, 20 million/mL, and 10 million/mL, with the 20 million/mL figure being the generally accepted standard. Confidence in any figure is limited by the inaccuracies inherent in the methods used for counting sperm. When a group of technicians and pathologists used a counting chamber to do a semen analysis on the same pooled specimen, the mean sperm count was 46.7 million/mL with a range of values lying between 10 and 98 million/mL, giving a coefficient of variation of 37.8%. These inconsistent results are not necessarily remedied with computer-assisted semen analysis (CASA). Significant errors can still occur with CASA, especially with specimens containing low numbers where other cells can be miscounted as sperm.

In an attempt to increase the prognostic value of the routine semen analysis, the results of the sperm count and the percentage of motility have been combined to give total motile sperm count (total sperm in the ejaculate x percent motility) or motile sperm per milliliter (sperm per mL x percent motility). The combining of count and motility allows determination of the number or concentration of active sperm and may provide a more informative way to present the data. Males with less than 24 months of infertility and a motile sperm count of greater than 10 million/mL have a higher pregnancy rate than those with 2–10 million/mL motile sperm. Even with motile sperm counts less than 5.1 million/mL, the pregnancy rate was 37.5%. However, with counts between 60.1 and 100 million motile sperm/mL, the pregnancy rate rose to 78.6%. Whereas trends are evident, the overlapping results limit the prognostic power of combining sperm count and sperm motility.

Sperm Morphology

Until recently, the evaluation of sperm morphology was beset with the same variations that plague sperm counts. Katz and Overstreet introduced an overlay to use with video microscopy, which allowed a more standardized assessment of morphology. With this method the coefficient of variation between observers was markedly reduced; 50% of sperm should have normal morphology when assessed with this technique.

Kruger and coworkers championed morphology as the best prognostic indicator for subsequent successful fertilization with in vitro fertilization. They utilized “strict criteria” that shift many sperm out of the normal category, by including as abnormal, sperm with even minor abnormalities as well as those with abnormalities of the acrosome. Using these strict criteria, males with greater than 14% normal forms had normal rates of fertilization with in vitro fertilization, whereas those with less than 4% normal forms had fertilization rates of only 7–8%. Values between 4% and 14% normal forms were associated with intermediate rates of fertilization. Many andrology laboratories have adopted “strict criteria” for the assessment of sperm morphology. In a dissenting study, the usual criteria for assessment of sperm morphology provided better prognostic information than did assessment by strict criteria. Technicians well trained in using strict criteria can provide highly reproducible results, but the standardization may not be possible on a more widespread scale. Interobserver differences in assessing sperm morphology could be decreased with the utilization of computer-assisted morphometric evaluations.

Other Parameters

Whereas the count, motility, and morphology of the specimen constitute the major parameters on which the male’s fertility is categorized, there are other characteristics of the semen that may impact on fertility potential. A volume of less than 1 mL may be too small to make contact with the cervix, and a volume greater than 7 mL may dilute the sperm concentration so that insufficient numbers are in close proximity to the cervix.

Round cells in the specimen can be either white cells or immature cells, and these look similar in wet preparations. The WHO standards manual states that a normal ejaculate should not contain more than 5 million round cells per mL (5 per high power field) while the number of leukocytes should not exceed 1 million per mL. There are staining methods, biochemical tests, and immunologic techniques to differentiate immature cells from white cells. In laboratories where these tests are not performed, all round cells are lumped together and reported as white blood cells. It is reasonable to obtain a culture, perhaps by prostatic massage, when the report shows that there are fewer white cells per high power field, even though some of these may be immature cells, and the presence of white cells does not always correlate with infection.

Repetitive agglutination of sperm (except when it is on pieces of debris) is suggestive of an immunologic effect or an infection. It can, however, be nonspecific and of no significance. Although it is common practice to evaluate the pH of semen because abnormalities may provide a clue to disorders of the accessory glands, in practice this measurement is of little value.

Tests of Sperm Function

Although all of the major elements of the semen analysis (numbers, motility, and morphology) have some bearing on fertility, especially when markedly deficient, the lack of precise correlations has led to a search for tests of the functional capacity of sperm. Despite enthusiasm generated by a variety of assays over the past 4 decades, no test has emerged as a reliable standard for the fertilizing ability of sperm. One problem is that the functional tests individually measure only one aspect of the many functions performed by sperm, whether it is attachment to the zona, penetration of the egg membrane, or the release of enzymes. Recently, there has been a trend in centers of assisted reproductive technologies to dispense with functional tests because the ultimate end point of fertilization of the human egg can now be tested by ICSI. If the functional tests were better prognosticators and less cumbersome to perform, they still could play a role in clinical practice. For example, a couple with unexplained infertility and an abnormal result might consider donor insemination as a therapeutic option.

Sperm Penetration Assay (SPA)

The zona pellucida of most mammalian species presents not only a block to polyspermy but also a barrier to fertilization of an egg by sperm of a different species. However, foreign sperm can fuse with and penetrate an egg if the zona is removed by gentle enzyme digestion. In the sperm penetration assay, eggs are collected from superovulated golden hamsters, the zona are removed by enzyme digestion, and the denuded eggs are cultured for 2–3 hours with human sperm that have been washed and incubated overnight in culture media. Presence of a swollen sperm head in the egg cytoplasm is evidence of successful penetration. Most laboratories report the percentage of eggs penetrated and compare this figure to the percent penetrated by a known fertile sperm specimen (some laboratories use the criterion of number of sperm penetrations per egg with 2 or more considered normal).

Whereas the concept of the SPA as a measure of sperm fertilizing ability is an attractive one, the practical aspects of the test have hindered its standardization. For example, the source of the albumin used as the protein supplement in the media can influence the result as can use of resuspended compared to wild-type sperm. Moreover, an individual’s results in the SPA can vary over time. In addition, different laboratories utilize different cutoff points for the lower limit of normal penetration with the most common points being 0, 10, 14, and 20%.

Equally important has been a continuing controversy over the prognostic value of the test. A meta-analysis concluded that the test was not of value. Other authors, however, have found correlations with eventual fertility. An SPA result of greater than 19% was associated with a pregnancy rate of 48%, whereas below 20% eggs penetrated was associated with a pregnancy rate of 20%. However, even with an SPA of 0%, the pregnancy rate in this series was 16%. Thus, failure of the sperm to penetrate the hamster egg is not an absolute indication that the sperm cannot penetrate the human egg. Because of this limitation of the SPA, attempts have been made to optimize the test with a goal of eliminating these false-negative results. Strategies to eliminate or to lower the number of false-negative tests include treatment of sperm with follicular fluid, test yolk buffer, calcium ionophore, miniaturizing the test, and adjusting the concentration of albumin or the ions in the culture media. With any of these maneuvers an SPA showing no or low penetration is a more accurate (although not 100%) indicator of poor sperm fertilizing ability. If an optimized SPA has zero penetration, the couple should be given the option of considering use of donor sperm or ICSI. In contrast to the problems with low SPAs, normal levels of sperm penetration correlate quite well, although not absolutely, with human fertilization in vivo and in vitro.

If there is a place in clinical practice for the SPA, it is to identify abnormalities of sperm not evident by studies of count, motility, or morphology. In cases of unexplained infertility, this could shorten the time period before pursuing either donor insemination or ICSI.
Human Zona Binding Assay

Whereas the SPA tests the ability of sperm to penetrate or to be engulfed by the egg, it does not test the critical ability to pass through the zona pellucida. The zona is, of course, removed in preparation for the SPA because it is, with rare exceptions, impervious to foreign sperm. Thus, to test zona penetrating or zona binding ability of human sperm requires the use of human zonae. One approach is to use zonae obtained from surgically removed ovarian tissue and slit them in half so that both patient sperm and donor sperm can be tested in parallel on different portions of the same zona. The ratio of the number of sperm bound for the test subject to the number of sperm bound for fertile control sperm has been labeled the hemizona assay index (HZI). A breakpoint at an HZI value of 36 has provided a good correlation with results in human IVF. The limited availability of zonae and the technical requirements of the assay will always restrict its application to a small number of committed laboratories. In the future, the development of materials that mimic the properties of the zona could lead to simpler tests. However, the widespread use of ICSI, which bypasses the zona, renders such tests superfluous, unless they can determine with certainty which couples require ICSI.

In Vitro Tests of Sperm Penetration into Mucus

Some of the vagaries of the postcoital test can be eliminated by use of in vitro testing with bovine cervical mucus or an artificial mucus. A drop of sperm is placed next to mucus on a slide and progression of sperm into the mucus monitored through a microscope. To better standardize the test, tubes filled with bovine cervical mucus, available commercially, can be utilized and the length of mucus traversed by the sperm measured. One problem is that antibody-affected sperm are not handicapped in moving through bovine mucus, whereas they usually move poorly in human cervical mucus. Because couples with poor postcoital tests and those with unexplained infertility will almost always be treated with intrauterine inseminations of washed sperm (IUI), with or without gonadotropins, there is little practical value for the in vitro sperm mucus tests in current practice.

Assessments of Sperm Motility

A sperm quality analyzer uses an electro-optical method to provide an assessment of the number of motile sperm in a specimen and, to some extent, the quality of the motility. This test provides some indication of the functional capacity of the sperm, but it has not gained widespread use.

Measurement of Sperm Velocity

The CASA systems are best at supplying information on sperm velocity and specific movements, such as lateral head displacement. However, it is unlikely that these measurements provide information that cannot be obtained with less expensive methods. The computer printouts of multiple measurements by CASA instruments provide a wealth of information, but little that is of current value for the clinician.

Hypo-osmotic Swelling Test

When sperm are placed in a hypo-osmotic solution of sodium citrate and fructose, a sperm tail with normal membrane function will swell and coil as fluid is transported across the membrane. Conversely, if there is a functional disturbance of the tail membrane, the tail will appear unaffected. This test has been scrutinized by a number of investigators with no clear-cut decision regarding its clinical usefulness. A specific niche has been suggested for the test. Non-motile sperm, but positive with the hypo-osmotic test, may provide better results with ICSI than those obtained with non-motile and hypo-osmotic negative sperm.

Measurement of Adenosine Triphosphate (ATP)

ATP is an important component of sperm metabolism related to tail movement. In one report, levels of ATP in semen were a strong discriminator between populations of fertile and infertile males. A multicenter study sponsored by the World Health Organization concluded, however, that levels of semen ATP could not predict the occurrence of pregnancy when the female was normal and the male partner had a sperm concentration greater than 20 million/ml.

Measurement of the Acrosome Reaction

The acrosome reaction that allows the release of enzymes from the acrosome (see Chapter 7) occurs on or near the zona pellucida. However, a low percentage of sperm will also become reactive while in media or following treatment with a calcium ionophore that induces capacitation. Although the artificial initiation of the acrosome reaction has been correlated with IVF results, the relatively small difference in acrosome-reactive sperm in the different groups indicates that this approach is not clinically important.

Measurement of Acrosin

Acrosin is a proteolytic enzyme associated with the acrosome that may be important for aiding sperm to traverse the zona. Low acrosin concentrations could be associated with infertility. Although theoretically appealing, the test has little application in clinical practice.

Sperm Antibodies

Whereas the previous assays measure sperm function or sperm numbers, sperm antibody tests determine reactions to sperm. It has been known for more than 100 years that animals, both male and female, can be rendered infertile by immunization with sperm. Semen is very antigenic and, fortunately, is normally isolated from the immune system by the blood-tests barrier. Disruption of this anatomic and functional barrier in the seminiferous tubules can lead to antibody formation; hence antibody production in the male can follow vasectomy, testicular torsion, infections, or trauma. In addition, there are women who have allergic reactions to semen manifested by reactions as diverse as irritation of the vagina and cardiovascular collapse following intercourse. The basic question is whether more subtle immunologic reactions can occur that interfere with fertility.

Initial efforts to detect sperm antibodies involved incubating sperm in the sera of both males and females with agglutination being the endpoint. Despite the fact that substantial agglutination of sperm in semen on a repetitive basis is an indication of the presence of antibodies, agglutination in serum often is nonspecific. Thus, testing with serum alone has been abandoned. It is now recognized that sperm antibodies in the circulation of men or women have no influence on fertility.

The two most widely used tests for sperm antibodies use immunologically mediated attachment of particles or beads to sperm that are assessed with a microscope. The immunobead test uses beads labeled with anti-IgG, anti-IgA, or anti-IgM, and, thus, it provides identification of the class of antibodies on the sperm. The site on the sperm where the beads are adherent also can be noted. Anti-IgG usually localizes to the tail and anti-IgA to the head of the sperm. Antibody localized only to the tip of the tail usually is not significant, whereas antibody on the rest of the tail may interfere with sperm motility. Antibodies on the head of the sperm can cause failure of fusion with the egg. A second test, the mixed agglutination test (SpermMar), uses antibodies to IgG to bridge antibody-coated sperm and latex particles that have been conjugated with human IgG. The endpoint in this test is clumping, and the reactions against individual segments of the sperm cannot be identified. The SpermMar test can be used on unprepared semen, as opposed to the immunobead test where sperm washing is required.

With the SpermMar test, the diagnosis of immunologic infertility is suggested when 10–39% of motile sperm are attached to the particles, whereas immunologic infertility is very probable when 40% or more of the motile sperm are covered with beads. Using the immunobead test, when over 50% of sperm were antibody-bound, the subsequent pregnancy rate was 15.3%, whereas when the percentage of sperm positive for antibody was less than 50%, the pregnancy rate was 66.7%. Others have used a cut off of 20%, and it is not clear at this time which is the more valid number.

An additional difficulty in diagnosing immunologic infertility is the fluctuations in antibody levels even without therapy. In males receiving a placebo, 58% had a decrease in sperm-bound immunoglobulins.

A high percentage of positive sperm antibody tests is associated with poor postcoital tests, and it would seem cost-effective to initially screen for antibodies only in individuals whose postcoital tests show no sperm, all dead sperm, a high percentage of shaking sperm, or less than 3 motile sperm per high power field. This latter...
number is somewhat arbitrary, and some may choose to use less than 5 or even less than 1 as the cut off value.

**Because both unexplained infertility and immunologic infertility are treated with intrauterine inseminations, the usefulness of sperm antibody testing is limited.**

**Treatment of Sperm Antibodies**

Use of condoms to avoid contact between sperm and the female with antibodies has been abandoned because of lack of efficacy. The current office treatments for sperm antibodies in the male are the use of steroids or, preferably, intrauterine inseminations with prepared specimens. In an alternative, but experimental, treatment, the sperm are separated on Percoll gradients, and then incubated with antibody beads. A population of sperm without antibody can be separated from the mix and utilized for insemination.

Moderate to high doses of corticosteroids have been used to treat sperm antibodies in the male. Reports of efficacy in reducing antibody levels and marginal increases in pregnancy rate have been balanced by sporadic reports of serious side effects such as aseptic necrosis of the femoral head and less severe side effects such as irritability. A more popular therapeutic approach is the intrauterine insemination of washed spermatozoa in conjunction with gonadotropin treatment (superovulation) of the female. Determination of the efficacy of this treatment has been hindered by difficulties in deciding what constitutes a positive sperm antibody test in the female and reports that lumped together patients who were antibody positive with others who may not have been afflicted with antibodies but who had poor postcoital tests. Nevertheless, the reported pregnancy rates warrant several cycles of IUI-supervolulation before considering the more expensive methods of assisted reproduction.

Use of in vitro fertilization is a reasonable final approach to the treatment of sperm antibodies. Success with in vitro fertilization is reduced in couples with sperm antibodies, but once fertilization occurs, the probability of pregnancy is not affected. The results with ICSI are very good, even in the presence of high levels of antisperm antibodies.

Donor insemination should be considered as an alternative to the more expensive methods of assisted reproduction.

**Investigation and Treatment of Male Infertility**

If the semen analysis is abnormal, inquiry should be made concerning the presence of the following factors, any of which can produce abnormal sperm quality and quantity.

1. History of testicular injury, surgery, or mumps.
2. Exposure to excessive heat. A small rise in scrotal temperature can adversely affect spermatogenesis, and a febrile illness can produce striking changes in sperm count and motility. The effect of the illness can be seen in the sperm count and motility even 2–3 months later. This reflects the 74 days required for a spermatozoon to be generated from a primary germ cell. Environmental sources of heat, such as the use of jockey shorts instead of boxer shorts, excessively hot baths, hot tubs, or occupations that require long hours of sitting; e.g., long distance truck driving, may all decrease fertility potential; however, none of these factors has ever been substantiated by clinical study.
3. Severe allergic reactions.
4. Exposure to radiation or to industrial or environmental toxins Those who believe there is a downward trend in sperm counts commonly place blame on environmental pollutants or chemicals, including chemicals that mimic the effects of estrogen. A study from Scandinavia did show lower sperm counts in males from an urban area compared to males in rural areas, suggesting an effect of urban pollutants. More direct evidence of a deleterious effect of environmental hazards is difficult to obtain because there is a reluctance of workers to produce the serial semen specimens that would be required for a thorough industrial study. In any case, the clinician should determine if a male with an abnormal semen specimen has had exposure to industrial or environmental toxins.
5. Heavy marijuana and alcohol use can depress sperm counts and testosterone levels. Marijuana inhibits the secretion of GnRH and can suppress reproductive function in both men and women, and cocaine use is known to reduce spermatogenesis. There is no evidence to indicate that moderate alcohol intake affects fertility. There is good evidence that fecundity is reduced in women who smoke, and this is probably true for men as well, although not all studies agree.
6. Coital frequency. Counts at the lower levels of the normal range may be depressed to below normal levels by ejaculations occurring daily or more frequently. Conversely, abstinence for 10–14 days or more to save up sperm may be counterproductive because the gain in numbers can be offset by the lower motility produced by the increased proportion of older sperm. For most couples, coitus every 36 hours around the time of ovulation will give the optimal chance for conception.

Certain drugs, including cimetidine, spironolactone, nitrofurans, sulfasalazine, erythromycin, tetracyclines, anabolic steroids, and chemotherapeutic agents, can depress sperm quantity and quality. Cephalosporins, penicillins, quinolones, and the combination of sulfamethoxazole and trimethoprim are relatively safe to use when there is concern about effects on sperm. Neurologic ejaculatory dysfunction can be caused by a-blockers, phenolamine, methyldopa, guanethidine, and reserpine. Resumption of spermatogenesis has been reported to occur within 2 years following discontinuation of anabolic steroids; however, it is not known whether all individuals will return to normal function.

6. Coital frequency. Counts at the lower levels of the normal range may be depressed to below normal levels by ejaculations occurring daily or more frequently. Conversely, abstinence for 10–14 days or more to save up sperm may be counterproductive because the gain in numbers can be offset by the lower motility produced by the increased proportion of older sperm. For most couples, coitus every 36 hours around the time of ovulation will give the optimal chance for pregnancy. However, studies in men with oligospermia fail to detect a decline in sperm numbers with sequential ejaculations, suggesting that limitations on coital frequency are not necessary.

7. Exposure to diethylstilbestrol in utero has been suggested, but not proven, as a cause of male infertility. Indeed, in the largest follow-up of men born to the women treated with diethylstilbestrol, no impairment of fertility or sexual function was detected.

**Urologic Evaluation**

In the presence of semen abnormalities, referral is made to an urologist in order to look for an anatomic abnormality, an infection, an endocrine or genetic disorder, or a varicocele. Evaluation by an urologist detects serious medical pathology (such as a tumor) in approximately 1 of every 100 patients.

**Anatomic Abnormalities**

Examination may reveal a physical impairment such as a marked hypospadias, which can cause sperm to be deposited outside the vagina. In rare cases of diabetes, with some neurologic diseases, or occasionally following prostatectomy or pelvic lymphadenectomy, there can be retrograde ejaculation into the bladder. Pregnancies have been reported after insemination of sperm obtained from alkalized urine or following treatment with a variety of drugs. Retrograde ejaculation may be only partial, and some men with this condition have small amounts of visible ejaculate.

Obstruction or absence of the vas deferens is a relatively uncommon cause of male infertility. If the ducts are congenitally absent, fructose which is produced in the seminal vesicles will be absent from the semen. Testicular biopsy can differentiate between a block in the outflow tract or primary damage to the testes. Testicular damage or maldevelopment can be found following mumps orchitis, cryptorchidism, or in association with Klinefelter's syndrome. Males with the latter genetic abnormality (XXY) usually have small testes and azoospermia. With blockage of the vas, sperm can be aspired from the epididymis and vasa efferentia. Successful fertilization in vitro can result in pregnancy.

It is important that any infection in the genitourinary tract, including those caused by mycoplasma and chlamydia, be treated even though the specific effect on male fertility is not established.

**Genetic Causes of Male Infertility**

There is a high prevalence of Y chromosome submicroscopic deletions in men with oligospermia. A cluster of genes that are deleted in men with azoospermia or...
oligospermia has been referred to as DAZ (“deleted in azoospermia”). Subsequently, at least 3 specific regions on Yq were identified and labeled as “azoospermia factors,” AZFa, AZFb, and AZFc (the same region as DAZ). From 7% to 10% of men with severe oligospermia have been reported to have deletions in the AZFb and AZFc regions on the long arm of the Y chromosome. These microdeletions correlate poorly with the numbers of sperm present in the semen analysis. These could be normal genetic variants and not the cause of the infertility; however, these deletions can now be transmitted with the use of ICSI, and it is now appropriate to consider Y chromosome deletion analysis prior to treatment with assisted reproduction. On the other hand, Y chromosome deletions are rare in men with normal semen findings. Indeed, the prevalence of overall chromosomal abnormalities is greater in men with abnormal semen than the prevalence of Y chromosome microdeletions. Although these genetic abnormalities are relatively infrequent, and there is uncertainty regarding the role of Y chromosome microdeletions, because of the consequences, there is strong support for genetic screening of all men considering ICSI. Indeed, 90% of couples receiving genetic counseling elect to undergo genetic screening.

Approximately 1–2% of infertile males have the congenital bilateral absence of the vas deferens. Most males with bilateral congenital absence of the vas deferens have at least two mutations in the cystic fibrosis transmembrane conductance regulator gene. Screening for cystic fibrosis mutations should be considered in males with bilateral absent vas deferens who are considering ICSI. A reasonable approach involves screening the female partner; the absence of cystic fibrosis mutations in the female reduces the risk of having a child with cystic fibrosis to a very low level.

Primary ciliary dyskinesia refers to a group of disorders with autosomal inheritance and caused by abnormal or absent motility in ciliary structures. Kartagener syndrome includes sperm immotility, bronchiectasis, sinusitis, and situs inversus.

The presence of azoospermia or severe oligospermia with an intact vas warrants karyotyping. The most frequent diagnosis is Klinefelter syndrome, which occurs in approximately 1 in every 500 males. Among other genetic abnormalities, azoospermia and hypogonadism can result because of a mutation that causes an abnormality in the beta-subunit of LH. An inactivating mutation in the gene for the FSH receptor is associated with varying degrees of suppression of spermatogenesis, indicating that FSH is not absolutely essential for sperm production.

Endocrine Disorders

Although endocrine disorders are an uncommon cause for infertility, testing for thyroid, gonadotropins, prolactin, and testosterone may occasionally uncover unsuspected abnormalities. FSH levels are elevated with germ cell aplasia, and testosterone levels are decreased in men who are hypogonadal. Hyperprolactinemia is commonly associated with impotence, and in the absence of impotence, measuring a prolactin level is unlikely to aid in the diagnosis.

Infusion of gonadotropin-releasing hormone (GnRH) can stimulate secretion of gonadotropins, and there have been occasional reports of the usefulness of this treatment as well as the administration of gonadotropin in males who have an isolated gonadotropin deficiency. Although nonspecific therapy with thyroid, clomiphene citrate, and human chorionic gonadotropin has been used extensively, there is no compelling evidence that it is beneficial. Clomiphene citrate can elevate the sperm count, but an associated increase in fertility does not occur.

A fundamental problem in most studies of the efficacy of drug therapy in male fertility is the lack of a control group for comparison. Investigators make the erroneous assumption that the spontaneous cure rate of male infertility is zero and that any pregnancy that occurs during or following treatment is due solely to that treatment. A number of studies, however, have attested to the spontaneous cure rate of male infertility. In one study approximately one-third of males, with counts below 10 million/mL, who were not treated, successfully impregnated their partners. In summary, hormone treatment of infertile males who do not have an endocrine disorder is almost always unrewarding, and it does not improve fertility beyond what occurs by chance.

Varicocele

A varicocele is an abnormal tortuosity and dilatation of the veins of the pampiniform plexus within the spermatic cord. Approximately 20% to 40% of infertile males, depending on the zeal of the search, have a varicocele, usually on the left side because of the direct insertion of the spermatic vein into the renal vein. Varicoceles, in all likelihood, exert their effects by raising testicular temperature, an effect mediated by increased arterial blood flow.

Approximately 10–15% of males in a general population have a varicocele on physical examination, but there is no evidence that males with normal semen characteristics need treatment even if a varicocele is present. They should be checked periodically, however, to be sure that there is no deterioration in their semen characteristics.

It has been difficult to perform randomized studies of varicocele repair. A trial in Melbourne, Australia, failed because of poor compliance. Because the authors told their patients that varicocele repair might not make a difference, only 283 of 651 men chose to have it done. In those who had the repair, the only impact on the semen analysis was an improvement in motility from 33.5% to 39.3%, the classically reported finding. The same change, however, was noted in the nonoperated group, and the pregnancy rates in both the operated and nonoperated groups were the same! A small randomized trial in Germany observed a significant increase in sperm concentration in the treated group, but the pregnancy rates were the same in both treated and non-treated groups. However, varicocele is more commonly found in men with abnormal semen, and there is evidence that a varicocele may exert an increasingly deleterious effect over time.

Ligation of varicoceles results in a 30–35% pregnancy rate. Although the beneficial effects of treatment of varicocele have been disputed by some investigators who found equal results without treatment, current clinical practice supports the utilization of varicocele ligation in those males who have infertility and an impaired semen specimen.

Larger varicoceles exert a greater effect than small ones. Very small varicoceles, diagnosed only by ultrasound, are not worth treating. Decreased size of the left testicle together with a varicocele is associated with a worse prognosis with treatment compared with cases in which the testicles are of normal size. Although surgical interruption of the internal spermatic vein is the usual treatment for clinically apparent varicoceles, there is also a nonsurgical approach that utilizes embolization to occlude the vein.

Reactive Oxygen Species

Increased levels of reactive oxygen species can cause damage to the sperm membrane. Substances such as peroxidase and hydrogen peroxide can be released by abnormal sperm and by white blood cells, and when elevated levels of leukocytes are present in the semen (with a negative culture), treatment with vitamin E and glutathione is advocated.

Intrauterine Insemination of Washed Sperm (IUI)

Inseminations of whole semen have a limited role in infertility. They are useful when, either because of physical or psychologic factors, it is not possible to deposit sperm in the vagina by intercourse. In addition they are obviously useful in donor insemination. In the past small amounts of untreated semen were used for infertility. Intrauterine insemination, but the potential for reactions to the proteins, prostaglandins, and bacteria in semen have made this approach a historical relic. In its place preparation is inseminated into the uterus. Because of the resuspension, the live sperm in the inseminate are accompanied by dead sperm and miscellaneous cellular
elements. In the alternative swim-up technique, the unagitated pellet and overlying medium are placed in an incubator at 37° for 30–60 minutes. This provides lower numbers of sperm in the medium portion compared to the resuspension technique, but it achieves a cleaner specimen and for this reason it is the method we prefer. However, with severely oligospermic specimens it may be necessary to use resuspension to obtain sufficient sperm for insemination.

Whereas all sperm preparation methods produce specimens with better motility and more uniform morphology, this improvement may not necessarily translate into increased pregnancy rates. When equal numbers of motile sperm separated from fresh and from poor specimens were used in the sperm penetration assay, sperm separated from the good specimens were superior in achieving penetration. Thus, there may be intrinsic defects in sperm from poor specimens that may affect even the best sperm from that cohort.

At least 1 million motile sperm should be inseminated because lower numbers are seldom associated with success. When more than 15 million motile sperm were inseminated, there was no increase in the pregnancy rate; however, there was an increase in multiple births when the inseminate exceeded 20 million motile sperm. In another study, pregnancy rates were substantially higher when more than 10 million motile sperm were inseminated compared with inseminations of less than 10 million. Others have also reported increased pregnancy rates when higher numbers of donor sperm were inseminated. Important factors in determining the final motile sperm count are the initial sperm count and sperm motility. Combining two ejaculates obtained 4 hours apart can increase the number of sperm available from oligospermic specimens.

Empiric therapy, consisting of clomiphene alone, gonadotropin alone, IUI, or IUI combined with clomiphene or gonadotropin, in the female, has increasingly been used for treatment of infertility of any origin. The greatest enthusiasm supports IUI combined with gonadotropin, which offers a number of possible advantages. It increases the number of oocytes that have the potential for fertilization and increases the number of sperm reaching the uterine cavity. It raises the woman's hormone levels and eliminates seminal plasma, which may contain factors inimical to sperm. The enthusiasm was fueled not only by these postulated advantages of gonadotropin-MIU but by a series of positive reports that appeared in the literature in the late 1980s and early 1990s. In a population selected for male factor infertility or poor postcoital tests, superovulation combined with IUI increased the monthly proportion of pregnancies approximately four times compared to that following IUI timed by the LH surge. The most striking advantage was seen in the subgroup of women who were treated with gonadotropin and HCG rather than clomiphene alone or clomiphene-gonadotropin/HCG combinations. Another study in couples with unexplained infertility demonstrated a pregnancy rate per cycle of 2.7% with IUI alone, 6.1% with gonadotropin alone, but 26.4% when gonadotropin/IUI was used. Very similar experiences have been reported by others (see Chapter 30).

Doddson and Haney surveyed the literature and found that gonadotropin and IUI fecundity was 8.7% for male factor and 17% for unexplained infertility. Their own experience, which was included in the survey, was 15% for each category.

Our current formulation concludes that IUI without associated superovulation has little value in the treatment of male infertility. Gonadotropin-IUI marginally increases the pregnancy rate to approximately 7–8% per cycle (from a baseline of 2–3% per cycle) when an abnormality is present in the semen. Several treatment cycles are warranted, except in the presence of severely compromised semen specimens, where an early resort to ICSI is indicated.

The timing of inseminations and the number of inseminations per cycle may influence the ultimate pregnancy rates. Most commonly IUI is timed for the day following the LH surge measured in the urine or at approximately 36 hours after an injection of human chorionic gonadotropin (HCG). Variations on these schedules can still be associated with reasonable pregnancy rates. Increased pregnancy rates have been obtained by doing two IUIs in a cycle, the day of and the day after the LH surge or the ovulatory injection of HCG. Others have reported no difference in results with one or two IUIs. Because of the lesser expense and the greater convenience, we perform one insemination for several cycles before considering multiple inseminations.

Prior down-regulation with GnRH agonist treatment does not enhance results with gonadotropin/IUI. Similarly, intraperitoneal or intrauterine inseminations of sperm, although sometimes attractive, have no proven advantage over intrauterine insemination. Moreover, intratubal transfer could increase the risk of infection. Infection with IUI is rare, probably in the range of 1 in 500 inseminations. Multiple pregnancies occur in approximately 20% of cases of gonadotropin/IUI, and the pregnancy loss rate is approximately 20%. Ovulation stimulation of the ovaries can be minimized, but not eliminated, by monitoring of ovarian follicle numbers and growth by ultrasound (reviewed in Chapter 30). The cycle should be canceled by withholding HCG if more than 3–4 follicles with diameters of 14 mm or greater are present when the largest follicle size (16 mm or larger) indicates the time for HCG injection. Estrogen monitoring is not usually required, as opposed to IVF cycles where stimulation is more aggressive.

Estrogen monitoring can be helpful if a lead follicle displays an abnormal rate of growth, either too fast or too slow. A value of less than 200 pg/mL per mature follicle raised the suspicion that a follicle does not contain a normal oocyte. If there are excessive numbers of small follicles (10–13 mm) associated with 3 or 4 leading follicles, a high estrogen level (over 2000 pg/mL) indicates an increased risk of hyperstimulation, and HCG administration should not occur.

Treatment of sperm with methylxanthines such as caffeine and pentoxifylline seem to enhance motility. The compounds inhibit cAMP phosphodiesterase which results in an increase in cyclic AMP. Despite some earlier enthusiasm, pentoxifylline offered no advantage in an IVF program, and a cautionary note was raised because the safety of the compound in early pregnancy has not been established.

Therapeutic Donor Insemination

The combined problems of male infertility and adequate availability of baby donors have increased the interest in, and the demand for, therapeutic donor inseminations (TDI). The procedure raises emotional, ethical, and legal questions that must be considered and discussed. The clinician must never do inseminations without the consent of both partners. Increasingly, single women are seeking TDI. McGuire and Alexander reported that children in single head of household families are as psychologically adjusted as those from two-parent households and that TDI should not be denied to single women solely on the basis of their lack of a marital partner.

Three points are worth emphasizing.

1. Donor inseminations do not guarantee pregnancy. In the past, the success rate with fresh semen was about 70% over 5–6 cycles. The current use of frozen semen lowers the success rate.

The fecundability (chance of getting pregnant per cycle) has been reported to be 18.9% with fresh semen and only 5.0% with frozen semen. However, with exceptionally good frozen specimens, success can approach that achieved with fresh specimens. Over 80% of pregnancies that will occur do so within 6 months with fresh semen and within 12 months with frozen semen. In a summary of nearly 3000 treatment cycles with frozen semen, the cumulative pregnancy rates were 21% at 3 months, 40% at 6 months, and 62% at 12 months for women less than 30 years old. For women over the age of 30, the pregnancy rates were 17%, 28%, and 44%, respectively. Because of the risk of acquired immunodeficiency syndrome (AIDS), use of frozen sperm that has been banked for 6 months is now accepted clinical practice. The donor is tested for HIV at the time of collection and again at 6 months. In highly selected cases, washed, swim-up sperm for intrauterine insemination from an HIV-infected man has been used to achieve a healthy pregnancy in his partner. Despite an absence of seroconversion of the female partner, the full safety of this approach has not been established. Of course, sperm prepared in this manner should not be used in a donor program.

2. The couple needs to give some thought to their feelings should the child be born with a congenital anomaly. This will occur in 4–5% of all pregnancies, irrespective of whether they follow intercourse or therapeutic donor insemination.

3. Both the woman and her partner, or the woman alone if there is no partner, should sign a consent form. The procedure is covered by law in many, but not all, states. It is worthwhile for the clinician to know the legal status of TDI in his or her state so that correct information can be conveyed to patients.

As a rule the donor should be unknown to the couple. However, there has been discussion of sealing personal information on the donor with the understanding that, if he gives permission, the information can be conveyed to children conceived with his sperm, when those children reach adulthood.

The health and fertility of the donor must be unremarkable, and there should be no family history of genetic diseases. Screening for thalassemia in Mediterranean races, Tay-Sachs heterozygosity in Jews, and sickle cell disease in blacks is a wise precaution. Donors can also be tested for cystic fibrosis. Potential donors who are at high risk for AIDS (homosexuals, bisexuals, intravenous drug users) should be excluded as should individuals who have multiple sexual partners. Similar exclusions include individuals with histories of hepatitis, chronic hepatitis, and venereal diseases. In addition, testing is recommended for human immunodeficiency virus (HIV), syphilis, hepatitis B antigen, antibody to hepatitis B, hepatitis C, gonorrhea, chlamydia, and cytomegalovirus. As noted above, an initial screening test for HIV, if negative, should be repeated after 6 months. If both results are negative, the semen, which should be cryopreserved and quarantined for the 6 months, can be used. The donor will not be a mirror image of the male partner, but an attempt should be made to match physical characteristics.
Most individuals undergoing TDI consider it a private matter and not subject to discussion with family and friends. If successful in achieving pregnancy, some individuals discuss the origins of the conception with their children, but most people prefer to leave it unsaid.

Use of friends or relatives as donors raises the potential for emotional problems in the future, although we have used a relative when it was requested by a stable, intelligent couple who understood the long-term implications. The related donor must be subjected to the same laboratory screening as all other donors. Requests to mix the partner's sperm with the donor's signifies that the couple may not have made the emotional adjustment to the thought of donor insemination. A partner's semen also may impair the donor's sperm, although this is in dispute.

Donor inseminations are useful in azospermia, severe oligospermia, or asthenospermia refractory to treatment. They also are useful for the rare woman who has a history of fetal loss due to Rh sensitization. In that case an Rh-negative donor would be used. Genetic diseases may, on occasion, be an indication for donor insemination.

The basal body temperature (BBT) change, the woman's perception of vaginal wetness, and ovulatory pain, if present, are useful guides for timing of inseminations. More precise timing can be accomplished by monitoring of the day of the LH surge with measurements of LH in urine with any of a number of commercially available kits. In our experience approximately 75% of women can successfully use the kit at home to identify their LH surge. Insemination is performed the day after the LH surge is identified. In more difficult cases, monitoring and treatment approaches utilize ultrasound to monitor preovulatory follicle growth and an injection of 5000 or 10,000 IU human chorionic gonadotropin when the dominant follicle reaches 18 mm or greater in diameter.

If the BBT alone is used, an attempt is made to inseminate on the date just before or two days before the temperature rise with the timing based on reviewing 2 months of charts and/or the day of maximal vaginal wetness.

IUI with donor inseminations produces higher pregnancy rates compared with intracervical insemination. However, the multiple pregnancy rate may be slightly higher. One IUI per cycle should be performed for two cycles, and then, if no success, increase to two inseminations per cycle. Double inseminations in a donor program increase the pregnancy rate and shorten the time required to achieve pregnancy. IUI should be performed the day after a positive test with the urinary LH kit, or approximately 36 hours after HCG administration. When two IUIs are performed, they should be timed the day of and the day after the LH kit tests positive, or approximately 18 and 42 hours after HCG administration. A minimum of 6 months of donor inseminations, with or without superovulation, should be completed prior to moving on to in vitro fertilization. In a healthy woman, 3–4 cycles can be concluded prior to hysterosalpingography.

The children born after donor insemination have outcomes comparable to the general population. Interestingly, approximately half of couples do and half do not tell their children of their origins. The divorce rate in families with children conceived with donor insemination is lower than the general rate.

Intracytoplasmic Sperm Injection (ICSI)

ICSI, the direct injection of a single spermatozoon or spermatozoan into the cytoplasm of an oocyte (described in Chapter 31), was first described in 1992. The results have been nothing short of amazing, with pregnancy rates equal to or slightly surpassing the normal fecundity rate. Pregnancy rates are independent of any semen analysis parameter (number, motility, morphology). All that is required is the DNA material, and this can be derived from a barely moving spermatozoon or in men with no spermatozoa or spermatozoans in the semen analysis from a spermatozoon (or spermatozoa) obtained by epididymal aspiration or from a biopsy of the seminiferous tubules.

Although ICSI avoids the natural process of sperm selection, the chromosomal abnormality rates in the offspring have been equal to the rate in the general population, with the exception of a slight increase in minor sex chromosome abnormalities (that could pass the male infertility problem to children), and, thus far, growth and development of the children have been normal.

Conclusion

For many years, the outlook for male infertility (with the exception of donor insemination) was very pessimistic. Today, in vitro fertilization with ICSI can overcome even the most grievous cases. The availability and expense of this technology, however, often are out of the reach of many couples. Therefore, less rigorous treatment continues to be an important part of the treatment of male infertility. It is important to identify a urologist in the local community who has both an interest and expertise in male infertility. Even without that resource, the clinician has the ability to offer IUI and superovulation.

Chapter References

15. MacLeod J, Gold RA. The male factor in fertility and infertility. II. Spermatozoa counts in 1000 cases of known fertility and 1000 cases of infertility marriage, J Urol 66:436, 1951.
31. Burkman LJ, Coddington CC, Franken DR, Kruger TF, Rosenwaks Z, Hodgson GD. The henirosis assay (HZA): development of a diagnostic test for the binding of human spermatozoa to

Copyright ©
Induction of Ovulation

In previous editions of this book, we began this chapter with a statement in celebration of one of the greatest achievements of reproductive endocrinology, the ability to induce ovulation and attain pregnancy in women who in the past had little basis or hope for reversal of their ovulatory dysfunction. As we approach a half century of ovulation induction for infertility, this elation remains justified. A variety of logical strategies, often empirically defined, designed to respond to specific indications do yield excellent results. In many clinical circumstances accurate data confirm the theoretical expectation that treatment can result in pregnancy rates equivalent to those in the normal population.

As the field has matured, however, the initial exuberance has been tempered by the realities of objective review. Results of treatment, despite restriction to a single indication, are not uniform. Complications occur, and escalation to more costly and hazardous treatments is a common occurrence. Perhaps most concerning is the vexing persistent disparity between rates of successful ovulation (high) and pregnancy (relatively low). This paradox presents two challenges.

1. The informed consent dialogue with the infertile couple does not conclude with the initial outline of strategic options. Helping the couple deal with the frustration of repeated partial success but persistent ultimate failure requires special time, sensitivity, and compassion.
2. The disparity between ovulation and pregnancy rates is an emphatic reminder for the clinician of the imprecision of the diagnosis, "ovulatory dysfunction and/or failure," the inadequacy of the available tests to determine if induction of ovulation has actually occurred, and the existence of more basic inherent defects in the reproductive process not disclosed by our current thorough work up or simply not discovered.

These considerations are all the more reason for the clinician to understand thoroughly the various indications and the many options available for induction of ovulation. This chapter will review the principles which guide the use of clomiphene, human menopausal gonadotropins, purified follicle-stimulating hormone (FSH), recombinant FSH, bromocriptine, and gonadotropin-releasing hormone (GnRH), and consider the results and complications of the medical induction of ovulation. In addition, laparoscopic ovarian surgical procedures (modified wedge resection) and the impact of GnRH agonist therapy will be reviewed.

Despite the specificity of the therapy and the promise of successful results, it is incumbent upon the practitioner to perform the appropriate medical evaluation to ensure that a contraindication to therapy is not overlooked. The reader is referred to Chapter 12 and Chapter 13 for a consideration of anovulation and hirsutism, and Chapter 11 for the evaluation of amenorrhea and galactorrhea.

For reference purposes, we provide the following definitions of ovulatory deficiencies according to the World Health Organization.

**Group I: Hypothalamic-Pituitary Failure**

This classification includes patients diagnosed as hypothalamic amenorrhea, and includes stress-related amenorrhea, anorexia nervosa and its variants, Kallmann's syndrome, and isolated gonadotropin deficiency. These patients display hypogonadotropic hypogonadism with low gonadotropin and estrogen levels, normal prolactin concentrations, and a failure to bleed after the administration of a progesterational agent (the progesterational challenge).

**Group II: Hypothalamic-Pituitary Dysfunction**

This classification includes normogonadotropic, normoestrogenic, anovulatory, oligoamenorrheic women. The classic anovulatory polycystic ovary syndrome is in this category.

**Group III: Ovarian Failure**

Patients in this classification are hypergonadotropic hypogonadal individuals with low estrogen levels. All variants of ovarian failure and ovarian resistance are in this category.

For the purposes of this chapter, as well as the expression of our clinical philosophy, hyperprolactinemic ovulatory dysfunction is treated as a specific treatment entity.

**Induction of Ovulation With 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. Clomiphene citrate is an orally active nonsteroidal agent distantly related to diethylstilbestrol. Clomiphene is a racemic mixture of its 2 stereochemical isomers, originally described as the cis and trans isomers. This designation is now recognized to have been inaccurate, and the isomers have been relabeled as zuclophenone and enclomiphene citrate. Clomiphene is available in 50 mg tablets, which contain 38% of the more active zuclophenone form.

The similarity of clomiphene's structure to an estrogenic substance is the clue to its mechanism of action. Clomiphene exerts only a very weak biologic estrogenic effect. The structural similarity to estrogen is sufficient to achieve uptake and binding by estrogen receptors; however, there are several important different characteristics. Perhaps most importantly, clomiphene occupies the nuclear receptor for long periods of time, for weeks rather than hours. Clomiphene modifies hypothalamic activity by affecting the concentration of the intracellular estrogen receptors. Specifically, the concentration of estrogen receptors is reduced by...
During clomiphene administration, circulating levels of FSH and LH rise. The subsequent ovulation that occurs after clomiphene therapy is a manifestation of the hormone and morphologic changes produced by the growing follicles. Clomiphene therapy does not directly stimulate ovulation, but it retrieves and magnifies the sequence of events that are the physiologic features of a normal cycle. The effectiveness of the drug, however, may not be restricted to its ability to cause an appropriate GnRH discharge.

In animal models, clomiphene exerts an estrogenic effect on the pituitary and directly stimulates gonadotropin release, independent of its action on GnRH. In the presence of estrogen, clomiphene influences pituitary response to GnRH in women, preferentially promoting FSH secretion. In addition, clomiphene exerts a direct ovarian effect. In the absence of estrogen, clomiphene stimulates LH receptors and LH release. In women with anovulation, clomiphene treatment does not result in ovulation, but it may result in the development of multiple follicles. Anovulatory women, however, respond in a different fashion. Clomiphene stimulates an increase in gonadotropin pulse amplitude, presumably because GnRH pulses are already operating at maximal frequency in anovulatory women with polycystic ovaries. Nevertheless, the experimental data indicate that the primary site of action is the hypothalamus.

The similarity of clomiphene’s structure to an estrogenic substance is the clue to its mechanism of action. Clomiphene exerts only a very weak biologic estrogenic effect. The structural similarity to estrogen is sufficient to achieve uptake and binding by estrogen receptors; however, there are several important different characteristics. Perhaps most importantly, clomiphene occupies the nuclear receptor for long periods of time, for weeks rather than hours. Clomiphene modifies hypothalamic activity by affecting the concentration of the intracellular estrogen receptors. Specifically, the concentration of estrogen receptors is reduced by inhibition of the process of receptor replenishment.

When exposed to clomiphene, the hypothalamic-pituitary axis is blind to the endogenous estrogen level in the circulation. Because receptor capacity is reduced and the true estrogen signal falsely lowered, negative feedback is diminished and the neuroendocrine mechanism for GnRH secretion is activated. When clomiphene is administered to normally cycling women, FSH and luteinizing hormone (LH) pulse frequency (but not amplitude) is increased, suggesting an increase in GnRH pulse frequency. Anovulatory women, however, respond in a different fashion. Clomiphene stimulates an increase in gonadotropin pulse amplitude, presumably because GnRH pulses are already operating at maximal frequency in anovulatory women with polycystic ovaries. Nevertheless, the experimental data indicate that the primary site of action is the hypothalamus.

Clomiphene has been found to cause disruptions of the organization of the uterine mesenchyme and tubal epithelium in human fetal reproductive tissue transplanted into immuno-isologous immunodeficient mice. Clomiphene is simple, safe, and cost-effective.

Selection of Patients for Clomiphene Treatment

Cases of ovarian failure are unresponsive to any form of ovulation induction. Therefore, the presence of ovarian tissue capable of responding to gonadotropins must be documented. This is only a problem in the patient with amenorrhea, since the presence of menstrual bleeding confirms the function (although perhaps limited) of the hypothalamic-pituitary-ovarian axis. The patient with amenorrhea who fails to produce a withdrawal bleed after a course of a progesterational agent (medroxyprogesterone acetate, 10 mg daily for 5 days) must be further evaluated. A case has been made by others for the usefulness of an ovarian biopsy, perhaps via the laparoscope, to establish the presence of competent ovarian tissue. It is our practice, however, to rely on the immunoresistance of gonadotropin levels and the response to a progestin, thus avoiding unnecessary surgical and anesthetic risks, to accurately rule out hypergonadotropic hypogonadism (ovarian failure). Attempts at medical induction of ovulation in these patients would be a waste of time and money.

The patients most likely to respond to clomiphene display some evidence of pituitary-ovarian activity as expressed in the biologic presence of estrogen (spontaneous or withdrawal menstrual bleeding). These are anovulatory women who have gonadotropin and estrogen production, but do not cycle, or women with inadequate luteal phases.

If the mechanism of an inadequate corpus luteum is inadequate FSH stimulation during the follicular phase, it makes sense to treat this condition with clomiphene, and a good response has been observed by ourselves and others. Two randomized trials comparing clomiphene to progesterone treatment for inadequate luteal phases demonstrated equal pregnancy rates with each treatment. Clomiphene does not prolong the luteal phase (as progesterone supplementation does). This is an important advantage, avoiding the anxiety and heightened monthly emotional response of infertile couples.

The patient who is deficient in gonadotropin secretion and, as a result, is hypoestrogenic, cannot be expected to respond to further lowering of the estrogen signal and thus should not respond to clomiphene. However, this principle is not completely applicable to clinical practice. An occasional patient who is, by all criteria, hypoestrogenic will respond. Therefore, any otherwise medically uncomplicated patient with infertility secondary to lack of ovulation is a candidate for clomiphene.
The patient who is deficient in gonadotropin secretion and, as a result, is hypoestrogenic, cannot be expected to respond to further lowering of the estrogen signal and thus should not respond to clomiphene. However, this principle is not completely applicable to clinical practice. An occasional patient who is, by all criteria, hypoestrogenic will respond. Therefore, any otherwise medically uncomplicated patient with infertility secondary to lack of ovulation is a candidate for clomiphene therapy unless galactorrhea or hyperprolactinemia is present. Hypoestrogenic women respond so rarely, however, that it is appropriate to omit treatment with clomiphene and move to other more productive options.

In addition to anovulation, treatment with clomiphene is indicated to improve the timing and frequency of ovulation and to enhance the possibilities of conception in the patient who ovulates only occasionally. Clomiphene can also be used to regulate the timing of ovulation in women undergoing insemination.

There is one special group in whom clomiphene is indicated in women who ovulate regularly and spontaneously. Certain religious requirements, such as those in Orthodox Judaism, interfere with the normal reproductive process. In the devout Orthodox Jewish couple, intercourse is prohibited in the presence of sexual flow and for 7 days following its conclusion. In some women menstrual flow is prolonged or the follicular phase is shortened, so that coitus cannot take place until after ovulation. In the usual mode of treatment, medication is begun on day 5 of the cycle. Ovulation can be delayed to a more appropriate time by starting clomiphene later, usually on day 7 or 8 of the cycle. Ovulation can be expected in the interval 5–10 days after the last day of medication. This manipulation has its limitations. Administration too late in the cycle, beyond day 9, may have no effect.

The question is often asked whether the indications for clomiphene therapy should be extended to include the initiation of cyclicity in the oligoamenorrheic patient who does not seek fertility. Increasingly inappropriately, women are being encouraged to consider treatment with clomiphene in the cycle in which it is used and it should not be expected to induce cyclicity following the conclusion of treatment, 2) the use of clomiphene may aggravate the clinical problems of acne and hirsutism during the treatment cycle by increasing LH stimulation of ovarian steroid production, and 3) the inability to induce cyclicity can be so discouraging to the patient that her acceptance of the drug will be impaired at some future date when it is legitimately offered as a fertility agent for the induction of ovulation.

Clomiphene and Unexplained Infertility

Clomiphene is used for the treatment of unexplained infertility: i.e., women who have prolonged (more than 3 years) infertility but who ovulate spontaneously and repeatedly by all available measures and do not have other abnormalities. While the spontaneous pregnancy rate is high in these patients (cumulative rates approach 50%), patient pressure and physician enthusiasm have led to superovulation induction in these couples. The rationale is appealingly clear. With more than one ovulation, surely there is an increased probability of successful fertilization, and such therapy would reverse (if present) episodic, unpredictable, recurrent, occult ovulatory dysfunction. Despite the high spontaneous pregnancy rate, clomiphene does have value in the empiric treatment of unexplained infertility, particularly prior to undertaking the more expensive and more complicated assisted reproductive technologies. In some studies, but not all, treatment with clomiphene has been associated with higher pregnancy rates. Superovulation for unexplained infertility is discussed in greater detail at the end of this chapter.

How to Use Clomiphene

A program of clomiphene therapy is begun on the 5th day of a cycle following either spontaneous or induced bleeding. It has not been established that a progestin withdrawal bleed is necessary before starting clomiphene treatment; we often omit this step if we are certain that the patient is not pregnant. The initial dose is 50 mg daily for 5 days. There is no advantage to beginning with a higher dose for the following two reasons: 1) in a random distribution of our patients begun with initial doses of either 50 mg or 100 mg daily, the pregnancy rate was identical; and 2) the highest incidence of side effects in our experience occurs at the 50 mg dose, and beginning with 100 mg, patients may develop more severe reactions. About 50% of patients conceive at the 50 mg dose, and another 20% at 100 mg. 

An occasional patient will be exceptionally sensitive to clomiphene and can achieve pregnancy at the reduced dose of 25 mg.

Beginning clomiphene on the 5th day is a method arrived at empirically; however, we can now offer a rational explanation based on ovarian physiology. The clomiphene-induced increase in gonadotropins during days 5–9 occurs at a time when the dominant follicle is being selected. Beginning clomiphene earlier can be expected to stimulate multiple follicular maturation resulting in a greater incidence of multiple gestation. Indeed, clomiphene is administered earlier in in vitro fertilization programs in order to obtain more than one oocyte. However, in standard ovulation induction protocols, no differences have been observed in the rates of ovulation, pregnancy, or spontaneous miscarriage whether clomiphene was started on day 2, 3, 4, or 5. 

If ovulation is not achieved in the very first cycle of treatment, dosage is increased to 100 mg. Thereafter, if ovulation and a normal luteal phase are not achieved in any cycle, dosage is increased in a staircase fashion by 50 mg increments to a maximum of 200–250 mg daily for 5 days. The highest dose is pursued for 3–4 months before considering the patient to be a clomiphene failure. The quantity of drug and the number of cycles go beyond those recommended by manufacturers. However, in our experience those recommendations are inappropriately limiting. We have achieved a 15% pregnancy rate at the 150 mg and 200 mg dose levels. 

There is a significant correlation between body weight and the dose of clomiphene required for ovulation. One must adhere to the usual regimen, however, because the weight cannot be used to predict prospectively the correct ovulatory dose. In other words, some obese women ovulate at the same low dose that achieves ovulation in thin women. Clomiphene is not stored in adipose tissue, and the increased dose often necessary in obese women is more likely due to a more intense anovulatory state with higher androgen levels producing a more resistant hypothalamic-pituitary-ovarian axis. Increasing the dose of clomiphene will eventually achieve the same level of success in overweight women as can be attained in lean women.

At the present time there is no clinical or laboratory parameter that can predict the dose of clomiphene necessary to achieve ovulation. Androgen and estrogen levels do not show any correlation with the dose of clomiphene that proves successful. Following the 5-day course of clomiphene, the ovulatory surge of gonadotropins can occur anywhere from 5 to 12 days after the last day of clomiphene administration, but most commonly on cycle day 16 or 17 when clomiphene is administered on days 5–9. Once ovulation is established, the subsequent treatment cycles are uniformly consistent. The patient is advised to have intercourse every other day for 1 week beginning 5 days after the last day of medication. In view of the role prostaglandins play in the physical expulsion of an oocyte, it is prudent to advise patients involved in programs of ovulation induction to avoid the use of agents that inhibit prostaglandin synthesis.

After the first treatment cycle the patient is evaluated for side effects, residual ovarian enlargement, and basal body temperature changes. We have found it unnecessary to perform a pelvic examination every month because significant ovarian enlargement is encountered infrequently, and it is usually symptomatic. It is more economical, for both patient and clinician, to mail the temperature chart to the office and several days later phone for the following month. Cysts within a reasonable size range (3–5 cm) do not require a rest from treatment; they do not respond to further stimulation, and they do not impact upon subsequent response.
Treatment of Hyperinsulinemia

After the first treatment cycle the patient is evaluated for side effects, residual ovarian enlargement, and basal body temperature changes. We have found it unnecessary to perform a pelvic examination every month because significant ovarian enlargement is encountered infrequently, and it is usually asymptomatic. It is more economical, for both patient and clinician, to mail the temperature chart to the office and several days later plan by telephone for the following month. Cysts within a reasonable size range (3–5 cm) do not require a rest from treatment; they do not respond to further stimulation, and they do not impact upon subsequent response.

We recommend basal body temperature charting to follow the response. If an inadequate luteal phase is evident, (temperature elevation less than 11 days duration) the amount of clomiphene is increased to the next dose level. If the patient is already at the maximal level, consider the available options for dealing with clomiphene failure (detailed below). Biophase changes are taken as an indication of ovulation and success. Maintenance of the temperature elevation beyond the expected time of menses is the earliest practical indication of pregnancy.

When the temperature chart is inconclusive or if the patient is not pregnant despite a period of apparently normal ovulations, an endometrial biopsy is indicated to document the adequacy of the luteal phase. One should also consider ultrasonographic monitoring for treatment failures, looking for follicles which do not reach mature size or the luteinized unruptured follicle syndrome (although it is by no means certain that luteinized unruptured follicles are a factor in infertility). For most patients, adding ultrasonography and urinary LH testing to basal body temperature charting does not improve pregnancy rates.

The additional use of HCG is limited to those cases in which there is a failure to ovulate at the maximal dose level or, when at that level, a short luteal phase is demonstrated. The rationale is to improve on the midcycle LH surge; therefore, 10,000 UI of HCG can be given as a single intramuscular dose on the 7th day after clomiphene when follicular maturation is at its peak. Because premature HCG administration may interfere with normal ovulation by down-regulating LH receptors, more accurate timing of HCG administration may be desirable. This requires either measurement of LH administration or estimation of follicular size (10–20 mm diameter) by sonography. When HCG is administered, intercourse is advised for that night and for the next 2 days. In our experience as well as in others reported in the literature, the addition of HCG has not had a significant impact on the pregnancy rate.

Care should be taken to review with the patient the pathophysiology of her condition, the principles of treatment, the prolonged course of therapy that may be necessary, and possible complications. Repeated failures accumulate frustration and despair in the couple, making each successive cycle of treatment more difficult. The anxiety and stress may hinder coital performance, and it is not uncommon for a couple to have difficulty performing scheduled intercourse.

Results With Clomiphene

In properly selected patients, 80% can be expected to ovulate, and approximately 40% become pregnant. The percent of pregnancies per induced ovulatory cycle is about 20–25%. The multiple pregnancy rate is approximately 10%, almost entirely twins. There have been rare cases of quintuplet and sextuplet births. In our own experience, with standardization of therapy, the incidence of twins has decreased.

The miscarriage rate is not increased. Most importantly, the incidence of congenital malformations is not increased, and infant survival and performance after delivery are no different from normal.

The discrepancy between ovulation rates and pregnancy rates is mainly due to 2 factors, the presence of other causes of infertility and a lack of persistence. In those patients with no other cause of infertility, the cumulative 6-month conception rate approaches a normal rate of 60–75%. The pregnancy rate per ovulatory cycle equals the normal rate. The pregnancy rate in 70% of our patients who received therapy sufficient to ovulate in at least 3 cycles was 55%, the same pregnancy rate after 3 months of exposure in the general population. With additional treatment cycles, the pregnancy rate decreases, although the ovulatory rate remains high. Approximately 15% of patients treated with the higher doses of 150–250 mg will become pregnant. Therefore there may be no need to attribute negative effects of clomiphene on ovocyte, endometrium, the corpus luteum, and an early embryo.

If all factors are corrected, and conception has not occurred in 6 months, prognosis is poor. In one large series, only 7.8% of those who had one or more infertility factors in addition to anovulation became pregnant.

Complications With Clomiphene

Side effects do not appear to be dose related, occurring more frequently at the 50 mg dose. Patients requiring the high doses are probably less sensitive to the drug. The most common problems are vasomotor flushes (10%), abdominal distension, bloating, pain, or soreness (5.5%), breast discomfort (2%), nausea and vomiting (2.2%), visual symptoms (1.5%), headache (1.3%), and dryness or loss of hair (0.3%). Patients who are extremely sensitive to the side effects of clomiphene can be successfully treated with half a tablet (25 mg) daily for 5 days, and even with half a tablet daily for 3 days.

A common antiestrogenic effect is an increase in the basal body temperature during the 5-day period of clomiphene administration. This is sometimes incorrectly interpreted as evidence of an early ovulation.

Visual symptoms include blurring vision, scotoma (visual spots or flashes), or abnormal perception. The cause of these symptoms is unknown, but in almost all cases, the visual symptoms have disappeared upon discontinuation of the medication. Usually these symptoms disappear within a few days but may take 1 or 2 weeks. There is one report of 3 women who had a persistence of visual side effects (shimmering, afterimages, and photophobia) after discontinuing treatment. It is likely that this problem is increased with increasing dosage and duration of exposure. Persistence of treatment in the presence of this reaction is not recommended.

Significant ovarian enlargement is associated with longer periods of treatment and is infrequent (5%) with the usual 5-day course. Maximal enlargement of the ovary usually occurs several days after discontinuing the clomiphene (in response to the increase in gonadotropins). If the patient is symptomatic, pelvic examination, intercourse, and undue physical exercise should be avoided because the enlarged ovaries are very fragile. Ovarian enlargement dissipates rapidly and only rarely is a subsequent treatment cycle delayed.

Because of multiple ovulations in some patients, it is possible that the risk of ectopic pregnancy is increased. However, evidence in the literature indicates that this increase at most is only slight.

What to Do With Clomiphene Failures

Patients most likely to not respond to clomiphene are those who are most hyperandrogenic and overweight (and probably insulin resistant). There are several options available for the 10–20% of women who fail to become pregnant with clomiphene up to the highest dose. Knowledge gained from experience in vitro fertilization provides us with several explanations for failure to respond to clomiphene. These include the effects of excessive LH in the follicular phase, the dysfunctional effects of an untimely LH surge, excess local concentrations of androgens, and hyperinsulinemia. These mechanisms may yield impaired folliculogenesis, increased atresia, poor oocyte quality, precocious or impaired oocyte maturation, low fertilization rates, variable implantation rates, and deficient corpus luteum function. Strategies have developed to mitigate or avoid many of these detrimental effects: the treatment of hyperinsulinemia, the supplemental use of dexamethasone (to reduce androgen burden), GnRH agonists (to eliminate endogenous LH intrusion), pulsatile GnRH therapy (to preserve physiologic interactive feedback mechanisms), and, finally, the use of human gonadotropins. We are also seeing the return of modified ovarian wedge resection by laparoscopic multiple ovarian tissue destruction with cautery or laser techniques.

First, make sure galactorrhea has not been overlooked and a prolactin level has been obtained. The good results with dopamine agonist treatment make it essential that this cause of anovulation be detected. It is worth performing a postcoital test to assess the quality of the cervical mucus. Despite the antiestrogen action of clomiphene the incidence of poor cervical mucus on the postcoital test is only 15%. In the past, estrogen (0.625 to 2.5 mg conjugated estrogens daily) was administered from day 10 to day 16 (for 1 week starting the day after last day of clomiphene administration) in an effort to improve mucus production. Although high doses of estrogen do not interfere with the gonadotropin response, ovulation, or the pregnancy rate, there is reason to believe that estrogen treatment is ineffective. Another alternative, the one we prefer, is to proceed with intrauterine inseminations of prepared sperm, bypassing the cervix.

After 6 months of clomiphene therapy, and in the absence of any other infertility factors, we proceed to one of the available options.

Treatment of Hyperinsulinemia

Anovulatory women with polycystic ovaries and hyperinsulinemia are more resistant to clomiphene treatment. The best therapy for these women who are obese is weight loss. Both the hyperinsulinemia and the hyperandrogenism can be reduced with weight loss, which is at least more than 5% of the initial weight.
Treatmet of Hyperinsulinemia

Anovulatory women with polycystic ovaries and hyperinsulinemia are more resistant to clomiphene treatment. The best therapy for these women who are obese is weight loss. Both the hyperinsulinemia and the hyperandrogenism can be reduced with weight loss, which is at least more than 5% of the initial weight. Other metabolic abnormalities are associated with an impressive rate of resumption of ovulation and pregnancy. In one study, 60 of 67 anovulatory women, who lost from 4 to 15 kg, resumed ovulation. A goal for weight loss that correlates with a good chance of achieving pregnancy, a reduction in insulin levels, and a decrease in free testosterone levels is a body mass index of less than 27 (see the nomogram for calculating body mass index in Chapter 11). It should be emphasized, that only a relatively small percentage of weight (5–10%) need be lost to have a beneficial impact upon insulin resistance and cardiovascular hemodynamic function.

It is reasonable to assume that all overweight, anovulatory women with polycystic ovaries are hyperinsulinemic. Nevertheless we recommend the measurement of the ratio of fasting glucose to fasting insulin in order to provide evidence that lends credence and importance to counseling efforts. A ratio of less than 4.5 is consistent with insulin resistance.

All anovulatory women who are hyperandrogenic should be assessed for insulin resistance and glucose tolerance with measurements of:

1. The fasting glucose:insulin ratio, followed by
2. The 2-hour glucose level after a 75 g glucose load:
   - normal
   - impaired
   - noninsulin-dependent diabetes mellitus

The best application of drugs, such as metformin and troglitazone, remains to be determined by data from appropriate clinical trials. Metformin improves insulin sensitivity, but the primary effect is a significant reduction in gluconeogenesis, thus decreasing hepatic glucose production. Metformin treatment (500 mg tid) reduces hyperinsulinemia, basal and stimulated LH levels, and free testosterone concentrations in overweight women with polycystic ovaries. A significant number of these anovulatory women treated with metformin ovulate and achieve pregnancy. In a group of obese women with polycystic ovaries, 90% of the women treated with metformin and 50 mg clomiphene ovulated compared with 8% in the group treated with placebo and clomiphene. However, there has been controversy, suggesting that the improvement was the result of the weight loss that often accompanies the use of metformin. In a study designed to control the effect of body weight, the administration of metformin was without effect on insulin resistance in extremely overweight women with polycystic ovaries. In another well-designed study, metformin again had no effect on insulin resistance when body weights remained unchanged, and in this study baseline weights and hyperinsulinemia were only modestly increased. In lean, anovulatory women with hyperinsulinemia, metformin treatment reduced hyperandrogenemia although there was no change in body weight; however, a decrease in the waist to hip ratio accompanied a reduction in the hyperinsulinemia. This study indicates that both obese and nonobese patients with hyperinsulinemia respond to metformin treatment. The reasons for the differences among the studies is not apparent. Perhaps only certain patients will respond to metformin, and, patient selection could influence the reported results. Lactic acidosis is a rare complication in patients treated with metformin; however, virtually all cases have occurred in patients with complicated medical problems, such as sepsis, renal insufficiency, and congestive heart failure. Patients who become ill should discontinue metformin treatment, and metformin should not be administered to women who have abnormal renal chemistries.

The thiazolidinediones markedly improve insulin sensitivity and insulin secretion (improved peripheral glucose utilization and b-cell function) without weight changes. Troglitazone (400 mg daily) decreases hyperinsulinemia, and improvements in metabolic abnormalities (decreased androgens, increased SHBG, decreased PAI-1 consistent with improved fibrinolytic capacity, and decreased LH), and a return to ovulation in very obese women have been reported with this agent. Although changes are common, liver function must be monitored during troglitazone treatment. The pharmaceutical company's package information should be consulted for the currently recommended monitoring regimen of serum alanine aminotransferase levels.

There is little doubt that these drugs can produce significant and beneficial improvements in this condition, although metformin may be effective only when weight loss occurs. However, is short-term use better than our standard methods of the induction of ovulation, and are these drugs safe during pregnancy and lactation? How effective are these agents in women who are of normal or only slightly elevated body weight? Appropriate clinical trials are required to answer these questions. In the meantime, clinicians have offered treatment with these agents to patients who have failed to respond to clomiphene. If an ovulatory response does not occur within 3 months, clomiphene treatment, beginning again at the lowest dose, is resumed in addition to the drug therapy for hyperinsulinemia.

The Addition of Dexamethasone to Clomiphene

Patients with hirsutism and high circulating androgen concentrations are more resistant to clomiphene. Dexamethasone, 0.5 mg at bedtime to blunt the nighttime peak of ACTH, is added to decrease the adrenal contribution to circulating androgens and, thus, diminish the androgen level in the microenvironment of the ovarian follicles. Higher ovulation and conception rates are achieved with this treatment when the circulating level of dehydroepiandrosterone sulfate (DHAS) is greater than the upper limit of normal. The dexamethasone is maintained daily until pregnancy is apparent. The dose of clomiphene is returned to the starting point of 50 mg and increased in incremental fashion as needed. One report indicated that even non-responders to clomiphene with normal DHAS levels achieved ovulation and pregnancy with the addition of a 10-day course of dexamethasone begun concurrently with clomiphene.

Extended Clomiphene Treatment

Several approaches have been reported for extending clomiphene treatment. We have very little experience with extended clomiphene treatment. In one approach, 250 mg of clomiphene was given for 8 days, followed by 10,000 IU HCG 6 days later. Three pregnancies were achieved out of 25 treatment cycles. In another series, the clomiphene dose was increased every 5 days, with some patients receiving up to 25 days of consecutive treatment, the last 5 days at 250 mg daily. Eight of 21 patients conceived and, in those patients who responded, measurement of gonadotropin revealed sustained elevations in FSH. This latter approach requires estrogen monitoring with discontinuation of the clomiphene when an increase in estrogen is detected. No patient ovulated after more than 21 days of treatment. Extending a dose of 100 mg clomiphene to 10 days has been reported to achieve pregnancies in women previously resistant to a 5-day regimen. A simple approach is to extend the duration of clomiphene treatment until a follicle of 18 mm diameter (on ultrasound) is obtained, then administer HCG.

Premtreatment Suppression

The anovulatory state is a dysfunctional condition. It is reasonable to expect suppression of the contributing factors to be followed by a reassertion of the harmony operating in a normal menstrual cycle, for a short time. Clinicians have long advocated (without benefit of careful study) a period of suppression (at least 6 months) prior to fertility treatment, followed by clomiphene, to reduce hyperandrogenemia. The same principle applies to the use of a GnRH agonist. Spontaneous resumption of ovulation has been reported to follow suppression for 6 months with a combination of a GnRH agonist and an oral contraceptive. Pretreatment suppression has added an attraction. It is reasonable to believe that women with elevated LH and testosterone levels not only have more difficulty achieving pregnancy, but also experience a higher spontaneous miscarriage rate. However, an excellent, randomized trial, testing suplatory suppression in women with polycystic ovaries, elevated LH levels, and recurrent miscarriages, obtained equal and good live birth rates with gonadotropin treatment in both the GnRH agonist-suppressed group and the non-suppressed group.

The Addition of Bromocriptine to Clomiphene

Although the use of bromocriptine to induce ovulation is indicated in the presence of galactorrhea or hyperprolactinemia, its use in the clomiphene failure patient with a normal prolactin and no galactorrhea is controversial. Anovulatory patients with normal levels of prolactin do respond to bromocriptine, but the effectiveness of this treatment has not been established by controlled studies. Nevertheless, the clinical response is occasionally impressive.

Bromocriptine. Elevated prolactin levels interfere with the normal function of the menstrual cycle by suppressing the pulsatile secretion of GnRH. This is manifested clinically by a spectrum, ranging from a subtle inadequate luteal phase to total suppression and hypoestrogenic amenorrhea. Regardless of the prolactin level, we interpret the presence of galactorrhea to indicate excessive prolactin stimulation. We screen all patients with galactorrhea or any ovulatory disorder with an assessment of the prolactin level. After a consideration of the problems of amenorrhea, galactorrhea, and the pituitary adenoma (as discussed in Chapter 11), bromocriptine emerges as the drug of choice for the induction of ovulation in these patients.

Ovarulatory dysfunction in the presence of galactorrhea responds well to bromocriptine, even if the prolactin level is normal. Either biologic activity is not being detected by the immunoassay or a random blood sample fails to reveal subtle elevations in prolactin.
Bromocriptine is examined in detail in Chapter 11, but it would be helpful to review pertinent details here. Bromocriptine is a dopamine agonist that directly inhibits pituitary secretion of prolactin. Suppression of prolactin levels restores CNS-pituitary gonadotropin function and also appears to increase ovarian responsiveness. The increase in ovarian responsiveness is seen in patients with normal prolactin levels and no galactorrhea. This is the apparent mechanism for an increase in sensitivity to clomiphene when bromocriptine is added to the therapeutic regimen. In women with persistent anovulation and polycystic ovaries, LH secretion is decreased by bromocriptine treatment, thus providing a rationale for why it might enhance the ovulatory response in these patients.  

The gastrointestinal and cardiovascular systems react to the dopaminergic action of bromocriptine, and, therefore, the side effects are mainly nausea, diarrhea, dizziness, headache, and fatigue. Side effects can be minimized by slowly building tolerance toward the usual dose, 2.5 mg bid. We start treatment with an initial dose of 2.5 mg at bedtime. If intolerance occurs, the tablet can be cut in half, and a slower program, developed by the patient, can be followed to work up to the standard dose. Usually, the second dose is added after 1 week, at breakfast or at lunch. In some patients, elevated prolactin levels can be reduced to normal levels with very small doses of bromocriptine, as little as 0.625 or 1.25 mg. Patients extremely sensitive to the side effects of bromocriptine can be treated by administering the drug intravaginally. Usually, one 2.5 mg tablet daily will be effective, if the prolactin level is elevated, the dose can be titrated to bring the prolactin level into the normal range.

The usual regimen is to administer bromocriptine daily until it is apparent the patient is pregnant, as usually determined by the basal body temperature chart. Although there has been no evidence of any harmful effects on the fetus, some patients and clinicians prefer to avoid bromocriptine in the luteal phase and, therefore, during early pregnancy. The drug is stopped when a temperature rise occurs and resumed when menses begin.

Ovulatory menses and pregnancy are achieved in 80% of patients with galactorrhea and hyperprolactinemia. Response is rapid, and, therefore, if there is no indication of ovulation (a rise in the basal body temperature) within 2 months, clomiphene is added to the regimen. The starting dose of clomiphene is 50 mg daily for 5 days, given and increased in the usual fashion.

Cabergoline is also an ergot-derived dopamine agonist. Patients resistant to both bromocriptine and quinagolide have been reported to respond to cabergoline. Cabergoline can be administered orally at doses of 0.5 to 3.0 mg once weekly, although it can be given twice weekly if necessary, with minimal dopamine agonist side effects (headache being the most common complaint). The low rate of side effects and the once-weekly dosage make cabergoline an attractive choice for the treatment of hyperprolactinemia. Bromocriptine can also be administered vaginally for the rare patient who cannot tolerate it orally.

**Bromocriptine for Euprolactinemic Women**
Clinical experience suggests that successful induction of ovulation and achievement of pregnancy with bromocriptine can occur in the absence of galactorrhea and with a normal prolactin level in women who have failed to respond to clomiphene. Some anovulatory women with normal prolactin levels who ovulated in response to bromocriptine have been found to have elevated nocturnal peaks of prolactin. The mechanism of action may be an increase in follicular responsiveness either due to suppression of prolactin or suppression of LH (a known action of dopamine). A decrease in LH may alter local follicular steroidogenesis in such a way to create a more favorable microenvironment. The method of administration is the same as above. If, after 2 months of treatment, there is no response, clomiphene is reinitiated, working up again from the starting dose of 50 mg daily. A carefully designed study has demonstrated that bromocriptine has not so far to offer for ovulatory women with unexplained infertility. On the other hand, bromocriptine or bromocriptine plus clomiphene treatment of ovulatory women with galactorrhea (and normal prolactin levels) yielded higher pregnancy rates when compared to a control group.

Once again the importance of detecting galactorrhea is emphasized.

**Ovarian Surgical Procedures**

The clomiphene-resistant patient can be treated by an ovarian surgical procedure. This subject is discussed near the end of this chapter.

**Induction of Ovulation With Human Gonadotropins**

For over 30 years, the only preparation used for gonadotropin treatment consisted of human menopausal gonadotropins, a preparation of gonadotropins extracted from the urine of postmenopausal women. The commercial preparation is still available, with either 75 units of FSH and 75 units of LH per ampule, or in an ampule with twice the amount, 150 units of each gonadotropin. Gonadotropins are inactive orally and, therefore, must be given parenterally; the heavy protein content of the urinary preparation requires intramuscular injections.

A more purified urinary preparation of FSH became available by removing most of the LH in the urinary product. This product still requires intramuscular injection. A more highly purified form is available that can be administered subcutaneously.

Recombinant FSH is now produced in Chinese hamster ovary cells transfected with the human FSH subunit genes. Recombinant FSH is homogeneous and free of contamination by proteins (characteristic of menopausal gonadotropins from urinary extracts); this allows simpler subcutaneous administration.

Because of its structural and biologic similarity to LH, HCG, readily available from human pregnancy urine and placental tissue, is used to simulate the midcycle LH ovulatory surge.

Recombinant HCG and recombinant LH are now available. These preparations will allow patient self-administration by subcutaneous injections. It is unlikely that the results will be different; however, it is possible that the risk of ovarian hyperstimulation will be reduced.

**PREPARATION**

Human menopausal gonadotropins
Purified urinary FSH
Highly purified urinary FSH
Recombinant FSH
Human chorionic gonadotropin
Recombinant HCG
Recombinant LH

**TRADE NAMES**

Pergonal, Humegon, Menogen, Repronex
Metrodin, Normegon, Orgafol
Fertexin or Metrodin HP
Puregon, Gonaf-F, Folistim
Pregnyl, Profasi, A.P.L.
Ovidrel
LHadi

**Selection of Patients for Gonadotropin Treatment**

Not only because of its expense but because of its greater complication rate, patients should not receive gonadotropin therapy without a very careful evaluation. An absolute requirement is the demonstration of ovarian competence. Abnormally high serum gonadotropins with a failure to demonstrate withdrawal bleeding indicate ovarian failure and preclude induction of ovulation except in those special cases of ovarian failure discussed in Chapter 11. Successful induction of ovulation and pregnancy has been reported in women with apparent ovarian failure, treated with a combination of estrogen (to suppress FSH to normal levels) and gonadotropins. Our own experience with this approach has been disappointing, and the chance of achieving pregnancy must be very low.

A thorough infertility investigation must be performed. In addition to the demonstration of ovarian competence, tubal and uterine pathology should be ruled out, anovulation documented, and semen analysis obtained. Nongynecologic endocrine problems must be treated. Hypogonadotropic function (low serum gonadotropins), including galactorrhea syndromes, requires evaluation for an intracranial lesion, with appropriate imaging and measurement of prolactin levels. It is imperative to take all steps necessary to exclude treatable pathology to which anovulation is secondary.

In our practice we sometimes, but not often, offer a course of clomiphene, not only because of the cost and complications associated with gonadotropin therapy, but also because some apparently hypergonadotropin patients will unpredictably respond to clomiphene. Because some patients cannot tolerate dopamine agonist treatment, it is important to know that hyperprolactinemia has no adverse effect on response to gonadotropins.

**How to Use Gonadotropin Therapy**
How to Use Gonadotropin Therapy

In our practice we sometimes, but not often, start a course of clomiphene, not only because of the cost and complications associated with gonadotropin therapy, but also because some apparently hypogonadotropic patients who predictably respond to clomiphene. Because some patients cannot tolerate dopamine agonist treatment, it is important to know that hyperprolactinemia has no adverse effect on response to gonadotropins. 

Ultrasound Monitoring

The guiding principle has been to administer HCG when mature follicles correlate with an estrogen level of 200–400 pg per mature follicle (greater than 13 mm diameter). This principle only applies when there are several leading follicles, not when many intermediate (9–16 mm) have been reported to predict poor response to gonadotropin stimulation. In this circumstance, perhaps a preparation containing some LH would be a better choice; for example, in patients with amenorrhea due to weight loss and anorexia.

A variable dosage method is used to achieve follicular growth and maturation. Follicle stimulation is achieved by 7–14 days of continuous gonadotropin, beginning with one ampule daily. Response is judged by the degree of estradiol produced by the growing follicles. The patient is monitored periodically with the measurement of the estradiol level and vaginal ultrasound assessment of the number and size of follicles. The patient is seen on the 7th day of treatment, and a decision is made to continue or increase the dose (step up method). After the seventh day, the patient is seen anywhere from daily to every one or two days. Another approach, the step down method, starts with a higher dose (2–3 ampules) and reduces the dose to one ampule after the initial response, theoretically approximating the changes in FSH in a normal ovulatory cycle.

Patients with polycystic ovaries are handled more gently because they are more sensitive to gonadotropin stimulation, and, therefore, at greater risk for multiple pregnancy and the ovarian hyperstimulation syndrome. The increased sensitivity to FSH is apparently due to the availability of a larger cohort of small follicles ready to respond to FSH (recruitable follicles). In these patients, monitoring begins on the 4th or 5th day of treatment. Excessive stimulation can be avoided in women with polycystic ovaries by using lower doses of gonadotropin extended over a longer duration of treatment. In addition, good results are inversely correlated with the degree of insulin resistance, and consideration should be given to treatment that improves insulin sensitivity (see elsewhere in this chapter and Chapter 12).

It cannot be emphasized too strongly that dosage administration and the judicious use of estrogen measurements depend upon the experience of the clinician administering gonadotropin therapy. When estradiol and ultrasound monitoring indicate that the patient is ready to receive the ovulatory stimulus, 10,000 units of HCG are given as a single dose intramuscularly. Neither manipulation of HCG dosage nor time of administration has been successful in changing the rates of multiple gestation and hyperstimulation. The patient is advised to have intercourse the day of the HCG injection and for the next 2 days. In view of the fragility of hyperstimulated ovaries, further intercourse as well as strenuous physical exercise should be avoided.

Pregnancy is usually achieved with the administration of gonadotropins for 7–12 days. The best results are obtained when the treatment period covers 10–15 days; when less than 10 days, the spontaneous miscarriage rate is increased. In general there is a direct relationship between dose and body weight; however, the same empiric approach is needed even in obese patients. In addition, individuals, presumably with extremely hypoosensitive ovaries, adequate follicular stimulation requires doses up to 4, 6, and more ampules/day. In this group of amenorrheic women massive doses of gonadotropins are necessary, and with proper monitoring, pregnancy can be achieved safely.

The range between the dose that does not induce ovulation and the dose that results in hyperstimulation is narrow. The situation is made even more difficult because the ovaries may react differently to essentially similar doses from month to month. Close supervision and experience in the use of gonadotropin therapy are necessary to avoid difficulties. There is no reason to avoid consecutive cycles of ovarian stimulation; indeed, an increased cycle fecundity has been observed in consecutive treatment cycles when compared to alternating stimulation and nontreatment.

Estrogen Monitoring

The use of estrogen measurements is necessary to choose the correct moment for administering the ovulatory dose of HCG in order to prevent hyperstimulation. On day 7 of the therapeutic cycle, blood is assayed for estradiol. Depending on the findings, the dosage of gonadotropin is individualized for the duration of the cycle. With experience, the clinician can avoid daily estrogen measurements, although sometimes this is necessary.

What should the blood estradiol level be? Because the blood estradiol is determined on a single sample of blood, the timing of the sampling with relationship to the previous injection of gonadotropin becomes a significant variable. When gonadotropin injections are given between 5 and 8 PM and blood samples are obtained first thing in the morning, an estradiol window of 1000–1500 pg/mL is optimal. The risk of hyperstimulation is significant from 1500 to 2000 pg/mL and, as a general rule, over 2000 pg/mL, HCG should not be given, and the ovarian follicles should be allowed to regress. Careful correlation of estrogen levels with the ultrasonographic picture allows a more aggressive approach. Haning has calculated that an upper limit for estradiol of 3800 pg/mL for anovulatory women (with polycystic ovaries) and 2400 for women with hypothalamic amenorrhea gives a risk of severe hyperstimulation of 5% in pregnant cycles and 1% in nonconception cycles.

Attempts to reproduce the normal midcycle increase of estradiol do not achieve a maximal pregnancy rate, and higher levels are required. The relative safety of this approach was seen in our series, where only 2 of 24 patients with estradiol levels over 1000 pg/mL developed hyperstimulation and it was moderate in both cases. When a higher dose of estrogen is used, timing of the HCG administration can be predicted fairly accurately by plotting the estradiol values on semilogarithmic paper. The rate of increase in estradiol is the same in spontaneous and induced cycles, and does not differ in cycles that result in multiple gestation. The level which is reached at the time of HCG administration is more critical than the slope of increase. Once a linear rise of estradiol is established, there is no need to increase the dose.

Ultrasound Monitoring

Ultrasound assessment of the growth and development of the ovarian follicle indicates the degree of follicular maturity and capability. During normal cycles, the growing cohort of follicles can be first identified by ultrasonography on days 5 to 7 as small sonolucent cysts. The dominant follicle will become apparent by days 15 to 18 mm in diameter as a fluid-filled cyst with a thickened wall. The follicle will grow rapidly during the last 24 hours of the cycle,

Ovulation is associated with complete emptying of the follicular contents in 1 to 45 minutes. Fluid can be detected frequently, but not always, in the cul-de-sac. The follicle either disappears or more commonly appears as a smaller, irregular cyst which diminishes in size over the next 4–5 days. In response to clomiphene treatment, follicles pursue a linear but generally accelerated rate of growth compared to spontaneous cycles. The maximal diameter of clomiphene-induced prevulatory follicles is similar to that seen with spontaneous cycles, 20–24 mm, but ovulation can be successfully induced with HCG administration when the diameter reaches 18–20 mm (by the time of ovulation, the follicle will have grown another 2–3 mm). With gonadotropin therapy, the maximal follicular diameter (15–18 mm) is smaller than that seen during spontaneous and clomiphene-induced cycles. When follicles reach this size, ovulation will occur approximately 36 hours after HCG is administered.

Ultrasound monitoring does not eliminate the risks of multiple gestation and hyperstimulation. It is claimed that a higher pregnancy rate can be achieved when ultrasound is combined with estrogen monitoring. The guiding principle has been to administer HCG when mature follicles correlate with an estrogen level of 200–400 pg per mature follicle (greater than 13 mm diameter). This principle only applies when there are several leading follicles, not when many intermediate (9–16 mm) are present. As a general rule, hyperstimulation is associated with the presence of more follicles. We believe that HCG should not be administered if there are more than 3–5 follicles 13 mm or greater in diameter (offering some protection against multiple gestation and hyperstimulation). Mild hyperstimulation has been associated with an increased number of intermediate size follicles, and severe hyperstimulation with an increase in small follicles.

A large number (11 or more) of smaller follicles also should preclude HCG administration.
The Hyperstimulation Syndrome

Ovarian hyperstimulation can be life-threatening. In mild cases the syndrome includes ovarian enlargement, abdominal distension, and weight gain. In severe cases, a critical condition develops with ascites, pleural effusion, electrolyte imbalance, and hypovolemia with hypotension and oliguria. 311 312 313 314 The ovaries are tremendously enlarged with multiple follicular cysts, stromal edema, and many corpora lutea. Because of this enlargement, torsion of the adnexa is a relatively common complication of this syndrome. 313

The incidence of clinically important hyperstimulation is striking. Although it might be expected that the mild type would be relatively common, the moderate to severe forms (grades II and III) are relatively rare. Two-fifths of cases occur early in the conception cycle, the remainder in nonconception cycles. For purposes of one of the methods of assisted reproduction, the ovaries are stimulated at an even greater rate than for conventional ovulation induction; however, the incidence of hyperstimulation has been associated with an increased number of intermediate size follicles, and severe hyperstimulation with an increase in small follicles. 313 314 A large number (11 or more) of smaller follicles also should preclude HCG administration.

Measurement of Endometrial Thickness. Ultrasonographic studies in cycles of in vitro fertilization have revealed that successful implantation is correlated with endometrial thickness on the day of HCG administration. 313 This correlation has been observed in patients in non-IVF patients receiving gonadotropins for ovulation induction. 313 Consistent with the antiestrogenic action of clomiphene, endometrial thickness is reduced in women treated with clomiphene (and reversed with estrogen). In a program utilizing clomiphene and timed administration of HCG for the purposes of intratubal inseminations, no pregnancies occurred when the endometrial thickness measured less than 6 mm. The chance of pregnancy is greatest, no matter what program of ovarian stimulation is being used, if endometrial thickness is 9–10 mm or more. 315 316

The Effect of Persistent Ovarian Cysts. In contrast to results with IVF or GIFT, the presence of a baseline ovarian cyst greater than 10 mm in diameter is associated with decreased fecundity in ovulation induction with gonadotropins. 315 With large cysts, treatment should be either delayed, or cyst suppression with a GnRH agonist or oral contraceptives should be considered, although it has never been proven that such treatment is effective. With this exception, there is no advantage in avoiding multiple successive treatment cycles (such as alternating treatment and nontreatment cycles). 315 316

Clomiphene-Gonadotropin Combination

The combination of clomiphene and gonadotropin was explored in order to minimize the amount and the cost of gonadotropin alone. As long as treatment is monitored with estrogen levels, the side effects and complications should not be dissimilar to those with gonadotropin alone. It has not been demonstrated that patients unresponsive to gonadotropin alone would respond to the sequence method, and there is no logical reason to assume that this would be true.

The usual method of treatment is to administer clomiphene 100 mg for 5–7 days, then to immediately proceed with gonadotropin beginning with 2 ampules per day. Estrogen levels are monitored as usual. The method may decrease the amount of gonadotropin required by approximately 50%; however, the same risks of multiple pregnancy and hyperstimulation can be expected. This reduced requirement for gonadotropin is found only in those patients who demonstrate a positive withdrawal bleeding following progestin medication or who have spontaneous menses. 313

Pulsatile Administration of Gonadotropin

Gonadotropin can be administered in pulsatile fashion, either subcutaneously or intravenously, using an appropriate pump system. 313 314 The aim is to reproduce the pulsatile pattern of gonadotropin secretion during the normal menstrual cycle. The dose administered intravenously is 8–9 units per pulse every 90 minutes, with adjustments to respond to the usual monitoring with estradiol levels and ultrasonography. It is not certain whether this complicated method is better beyond a decrease in dose and possibly a response in patients unresponsive to the traditional intramuscular regimen.

Results With Gonadotropin Treatment

The most significant aspect of this method of treatment is that it does achieve pregnancy in otherwise untreated situations. A cumulative conception rate of 90% after 6 treatment cycles can be achieved in women with hypothyroid amenorrhea (this rate exceeds that observed in spontaneously ovulating women), with a 23% rate of spontaneous miscarriage. 315 316 317 Women with normogonadotropic anovulation achieve only a slightly lower cumulative conception rate with relatively, but significantly; higher rates of miscarriage; however, the concomitant use of clomiphene and more careful gonadotropin treatment, these rates equal those in the normal population. 315 A slightly higher rate of spontaneous miscarriage reflects the combination of better detection of early pregnancy loss, advanced maternal age, and the increased incidence of multiple pregnancies. As with clomiphene, there is a normal incidence of congenital malformations, and the children have a normal postnatal development. 318 The risk of ectopic pregnancy is increased with ovulation induction, a consequence of multiple oocytes and high hormone levels. 313 These patients should be closely monitored in the early weeks of their pregnancies.

Women with polycystic ovaries and moderate obesity require larger doses of gonadotropin and ovulate at a lesser rate compared to leaner women with polycystic ovaries. 313 A comparable pregnancy rate (40%) is achieved. However, spontaneous miscarriage is more frequent, which is consistent with the observation that spontaneous miscarriage is more frequent in obese women.

HCG disappears from the blood with an initial component having a half-life of about 6 hours and a second, slower, component with a half-life of about 24 hours. It is this relatively slow half-life that enables a single injection of 10,000 IU to maintain the corpus luteum until pregnancy takes over. The HCG concentration after the ovulation injection should be less than 50–100 IU/L by day 14 after the injection. A subunit assay of HCG at this time or one of the urine assays performed 2–4 weeks after the HCG injection are reliable tests for pregnancy. Additional HCG (during the luteal phase) does not improve pregnancy rates. 313 Luteal supplementation is required only with concurrent treatment with a GnRH agonist.

The rate of serious hyperstimulation is 1–2%. Prior to the present era of more careful monitoring, the multiple pregnancy rate was reported as approximately 30% (triplets or more, 5%). Currently, the multiple pregnancy rate can be as low as 10% with careful monitoring and good medical judgment; however, rates as high as 40% are reported. 313 A striking increase in multiple births in many countries, especially triplets and quadruplets, can be directly attributed to ovulation stimulation for the treatment of infertility. 313 314 315

The multiple pregnancies are secondary to multiple oovulations, and, therefore, the siblings are not identical. The rate of spontaneous occurrence of twins is only about 1% and that of triplets 0.010–0.017% of the pregnant population. Dizygotic twinning varies among different populations and is inherited through the mother. The monogygotic twinning rate is about 0.3–0.4%, fairly constant, and uninfluenced by heredity. Surprisingly, induction of ovulation increases the frequency of monogygotic twinning 3–fold. 315 It is not known whether the multiple pregnancy rate with gonadotropin is significantly affected by a maternal history of twinning.

Maternal complications and fetal loss caused by prematurity in the multiple pregnancies have been serious problems. In addition, the miscarriage rate with gonadotropin is somewhat higher (25%) than normal, probably a combination of the effect of age, multiple pregnancies, and recognition of early miscarriages. 313

After at least one gonadotropin-induced pregnancy, the subsequent spontaneous pregnancy rate reaches 30% after 5 years. 313 Most of the pregnancies occur within 3 years of the gonadotropin-induced pregnancy, and the more endogenous hypothalamic-pituitary-ovarian function a patient has, the more likely a spontaneous pregnancy will occur. 313 This is consistent with a return to normal function in some women after suppression of a dysfunctional state. With time, it is logical to expect the original dysfunctional state to reestablish itself.

Multifetal Pregnancy Reduction. Induced abortion in the case of triplets or more is an option, but it would be surprising if patient and clinician would choose this solution. On the other hand, selective reduction of embryos in multiple pregnancy can be accomplished. 313 Under ultrasound guidance, a gestational sac can be aspirated or a cardiotoxic drug (potassium chloride) can be injected into, or adjacent to, the fetal heart. The transvaginal procedure is best performed between the 8th and 9th weeks of gestation and the transabdominal procedure between the 11th and 12th weeks. A later procedure is worthwhile because the spontaneous disappearance of one or more gestational sacs in multiple gestations can occur, an incidence of approximately 5% after fetal heartbeats have been identified. 313 315 Selection of which gestational sac to terminate is based solely on technical considerations, such as accessibility. The subsequent risk of losing one or more of the remaining fetuses is 4–5%, and of losing the pregnancy, 10% by experienced clinicians and higher with less experience. Reduction of a monozygotic pregnancy is not advisable because of shared vasculature and the high risk of losing all fetuses. The moral and ethical aspects of fetal reduction are significant, but in view of the potential problems associated with a multiple birth, it is a reasonable alternative for some.

The Hyperstimulation Syndrome

Ovarian hyperstimulation can be life-threatening. In mild cases the syndrome includes ovarian enlargement, abdominal distension, and weight gain. In severe cases, a critical condition develops with ascites, pleural effusion, electrolyte imbalance, and hypovolemia with hypotension and oliguria. 311 312 313 The ovaries are tremendously enlarged with multiple follicular cysts, stromal edema, and many corpora lutea. Because of this enlargement, torsion of the adnexa is a relatively common complication of this syndrome. 313

The incidence of clinically important hyperstimulation is striking. Although it might be expected that the mild type would be relatively common, the moderate to severe form (grades II and III) are relatively rare. Two-fifths of cases occur early in the conception cycle, the remainder in nonconception cycles. For purposes of one of the methods of assisted reproduction, the ovaries are stimulated at an even greater rate than for conventional ovulation induction; however, the incidence of

200–400 IU per mature follicle (greater than 10 mm diameter). This principle only applies when there are several leading follicles. For many women intermediate (9–16 mm) and small (6–8 mm) follicles are present. As a general rule, hyperstimulation is associated with the presence of more follicles. We believe that HCG should not be administered if there are more than 3–5 follicles 13 mm or greater in diameter (offering some protection against multiple gestation and hyperstimulation). Mild hyperstimulation has been associated with an increased number of intermediate size follicles, and severe hyperstimulation with an increase in small follicles. 313 314 A large number (11 or more) of smaller follicles also should preclude HCG administration.
The incidence of clinically important hyperstimulation is striking. Although it might be expected that the mild type would be relatively common, the moderate to severe form appears at an impressive rate (1–2%). Two-thirds of cases occur early in a conception cycle, the remainder in nonconception cycles. For purposes of one of the methods of assisted reproduction, the ovaries are stimulated at an even greater rate than for conventional ovulation induction; however, the incidence of hyperstimulation is no greater, or even lower. For this reason, it has been suggested that follicular aspiration offers partial protection against the hyperstimulation syndrome. Some reports have indicated a higher incidence of hyperstimulation in in vitro fertilization protocols using the combination of gonadotropin and a Gn-RH agonist. Because support during the luteal phase is required when Gn-RH agonist treatment is used, the additional HCG administered is responsible for the incidence and severity of hyperstimulation. The use of progesterone for luteal support is therefore, when estradiol levels are high (greater than 2500 pg/mL) and when the follicle number is greater than 15, progesterone supplementation is recommended instead of HCG.

Anovulatory women with polycystic ovaries are at greatest risk for the hyperstimulation syndrome. The use of a Gn-RH agonist in combined therapy does not eliminate this risk. These patients should be treated slowly with low titration of dose as described earlier.

The basic disturbance in hyperstimulation is a shift of fluid from the intravascular space into the abdominal cavity, creating a massive third space. The resulting hypovolemia leads to circulatory and electrolyte problems. The genesis of the ascites is unclear. The very high level of estrogen secretion by the ovaries may be the primary factor, inducing increased local capillary permeability and the edema of the ovaries. The volume of ascites is also critically related to the mass, volume, and surface area of the ovaries. Therefore, the larger the ovaries and the greater the steroid production, the more severe the condition. Or the syndrome may result from overproduction of autocrine/paracrine factors that affect vascular permeability.

A growing body of evidence implicates vascular endothelial growth factor (VEGF) in the pathophysiology of the hyperstimulation syndrome. The origin of VEGF is presumed to be the ovarian follicle, and increased capillary permeability and the severity of the syndrome are correlated with circulating levels of VEGF. Other cytokines, especially the interleukin family, are also believed to be involved in the permeability changes, and it is hypothesized that these agents affect the nitric oxide system. Large amounts of angiotensin II are present in the ascites produced by ovarian hyperstimulation, presumably originating from HCG stimulation of the ovarian renin-angiotensin system.

The loss of fluid and protein into the abdominal cavity accounts for the hypovolemia and hemococoncentration. This in turn results in low blood pressure and decreased central venous pressure. The major clinical complications are increased coagulability and decreased renal perfusion. Blood loss as the cause of the clinical picture can be easily ruled out since a hematocrit will reveal hemococoncentration. The decreased renal perfusion leads to increased salt and water reabsorption in the proximal tubule, producing oliguria and low urine sodium excretion. With less sodium being presented to the distal tubule, there is a decrease in the exchange of hydrogen and potassium for sodium, resulting in hyperkalemic acidosis. A rise in the blood urea nitrogen (BUN) is due to decreased perfusion and increased urea reabsorption and is not a sign of diuresis. Renal blood flow decreases as the volume of ascites increases; therefore, the level of sodium does not increase as much as the BUN. Thus, the patient is hypovolemic, azotemic, and hyperkalemic. In response to these changes, aldosterone, plasma renin activity, and antidiuretic hormone levels are all elevated.

Treatment is conservative and empirical. When a patient displays excessive weight gain (usually 10 or more pounds), excessive pain, hemococoncentration (hematocrit over 50%, while blood count over 25,000), oliguria, dyspnea, or postural hypotension, she should be hospitalized. Pelvic and abdominal examinations are contraindicated in view of the extreme fragility of the enlarged ovaries. Ovarian rupture and hemoperitoneum are easily precipitated.

Upon admission, the patient is put on bed rest, with daily body weights, strict monitoring of intake and output, and frequent vital signs. Serial studies of the following additional parameters: BUN, hematocrit, electrolytes, total protein, albumin/globulin ratio, creatinine, and urinary sodium and potassium. The electrocardiogram is utilized to follow and evaluate hyperkalemia. Fluid and salt restriction is controversial. It is argued that correction of the decreased circulating volume is not necessary as long as the BUN remains stable (an abnormally low urine output can be tolerated). Others believe that plasma expanding agents and electrolyte replacement should be administered. Human albumin (safety with concomitant infection) is the volume expander of choice. Potassium exchange resins may be necessary. Diuretics are without effect and, indeed, may be disadvantageous. The fluid in the abdominal cavity is not responsive to diuretic treatment, and diuresis may further contract the intravascular volume and produce hyperkalemic shock or thrombosis. Arterial and venous thromboses have been reported in both the upper and lower portions of the body, and anticoagulant therapy should be considered in severely hemococoncentrated patients. Although both antihistamines and indomethacin have been demonstrated to ameliorate the hyperstimulation in animal studies, their efficacy and safety in early human pregnancy are unknown. Inhibition of prostaglandin synthesis can worsen renal perfusion.

In severe cases, life-threatening adult respiratory distress syndrome can occur. These patients require intensive care monitoring, including central venous and pulmonary artery catheters, urinary bladder catheter, BUN, creatinine, electrolytes, total protein, albumin/globulin ratio, coagulation studies, and urinary sodium and potassium. The key point is that the hyperstimulation syndrome will undergo gradual resolution with time. In a patient who is not pregnant, the syndrome will cover a period of approximately 7 days. In a patient who is pregnant and in whom the ovaries are restimulated by the emerging endogenous HCG production, the syndrome will last 10–20 days.

The syndrome will not develop unless the ovulatory dose of HCG is given. Thus, the major emphasis in recent years has been to utilize monitoring to avoid hyperstimulation. Adult respiratory distress syndrome is associated with a 50% mortality rate. Chlorpheniramine maleate, an H-1 receptor blocker, appears to maintain membrane stability, allowing the use of fluids and mannitol to retain intravascular volume. With ultrasound guidance to avoid the enlarged ovaries, abdominal paracentesis can relieve severe pulmonary compromise. Aspiration of ascites can also be accomplished (and probably more easily) transvaginally (with ultrasound guidance). With repeated aspirations, it is very important to replace the lost plasma proteins (transfusion of the ascitic fluid has been used in a small number of cases). On the average, repeat aspirations are necessary in 3 to 5 days.

In severe cases, transvaginal aspiration of follicular structures should be considered to interfere with the intravascular mechanism responsible for the clinical picture; progesterone supplementation will be necessary to maintain an on-going pregnancy.

The possibility of ovarian rupture should always be considered, and serial hematocrits may be the only clue to intraperitoneal hemorrhage. Of course, a falling hematocrit accompanied by diuresis is an indication of resolution, not hemorrhage. Laparotomy should be avoided in these precarious patients. If surgery is necessary, only hemostatic measures should be undertaken and the ovaries should be conserved if possible, since a return to normal size is inevitable. If torsion of the ovary is suspected, acute management of the hemorrhage should be undertaken. If the ovarian hemorrhage is not controlled in time, a total hysterectomy may be necessary.

The key point is that the hyperstimulation syndrome will undergo gradual resolution with time.

What to Do With Gonadotropin Failures?

At one time, management of the couple who had failed ovulation induction with gonadotropin treatment was difficult, but straightforward. Short of another costly round if funds and emotional reserves were sufficient, nothing was left to offer except adoption. Today, a major option is now available in the assisted reproductive technologies (Chapter 31). Although even more costly and emotionally charged, these methods do offer significant additional opportunities for unsuccessful patients. Nevertheless, guidance to adoption services and emotional support continue to be part of the physician's obligation.

A very important question for patients is: when to stop? All clinicians have been confronted by the dismal and frustrating prospects of a woman who has squandered many corpora lutea, because of this enlargement, torsion of the adnexa is a relatively common complication of this syndrome.

What to Do With Gonadotropin Failures?

Women who present with "incipient" ovarian failure have elevated FSH levels and decreased levels of inhibin, but normal levels of estradiol. 140, 141 The rise in FSH during the later years is in response to declining inhibin production. 140, 142 and 143 Inhibin levels are lower in the follicular phase in women 45–49 years old compared with younger women. The specific inhibin involved in this feedback effect on FSH during the follicular phase is inhibin B. In addition, a decline in luteal phase inhibin A can contribute to a greater rise in FSH during the luteal-follicular transition. This decline in inhibin begins early, but accelerates after 40 years of age. The rise in FSH after age 40 is thought to be not a perfect one. Hyperstimulation has been found with relatively low estradiol levels, and high estrogen is not necessarily followed by hyperstimulation. As a general rule, the more follicles present (on ultrasound examination) the greater the risk for hyperstimulation. But this too is not a perfect correlation. Nevertheless, monitoring is the major available deterrent to a potentially life-threatening situation.

Can Failure to Respond Be Predicted?

Women who present with "incipient" ovarian failure have elevated FSH levels and decreased levels of inhibin, but normal levels of estradiol. The rise in FSH during the later years is in response to declining inhibin production. Inhibin levels are lower in the follicular phase in women 45–49 years old compared with younger women. The specific inhibin involved in this feedback effect on FSH during the follicular phase is inhibin B. In addition, a decline in luteal phase inhibin A can contribute to a greater rise in FSH during the luteal-follicular transition. This decline in inhibin begins early, but accelerates after 40 years of age. The rise in FSH after age 40 is thought to be not a perfect one. Hyperstimulation has been found with relatively low estradiol levels, and high estrogen is not necessarily followed by hyperstimulation. As a general rule, the more follicles present (on ultrasound examination) the greater the risk for hyperstimulation. But this too is not a perfect correlation. Nevertheless, monitoring is the major available deterrent to a potentially life-threatening situation.

Can Failure to Respond Be Predicted?

Women who present with "incipient" ovarian failure have elevated FSH levels and decreased levels of inhibin, but normal levels of estradiol. 140, 141 The rise in FSH during the later years is in response to declining inhibin production. 140, 142 and 143 Inhibin levels are lower in the follicular phase in women 45–49 years old compared with younger women. The specific inhibin involved in this feedback effect on FSH during the follicular phase is inhibin B. 140, 144 In addition, a decline in luteal phase inhibin A can contribute to a greater rise in FSH during the luteal-follicular transition. This decline in inhibin begins early, but accelerates after 40 years of age. The rise in FSH after age 40 is thought to be not a perfect one. Hyperstimulation has been found with relatively low estradiol levels, and high estrogen is not necessarily followed by hyperstimulation. As a general rule, the more follicles present (on ultrasound examination) the greater the risk for hyperstimulation. But this too is not a perfect correlation. Nevertheless, monitoring is the major available deterrent to a potentially life-threatening situation.

Can Failure to Respond Be Predicted?
inhibin A can contribute to a greater rise in FSH during the luteal-follicular transition. This decline in inhibin begins early, but accelerates after 40 years of age. The rise in FSH is very apparent after age 40, and there is no change in LH levels until menopause. 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). Arguing in favor of a role for a reduced follicular pool is the observation that follicular fluid obtained from prevulatory follicles of older women contains amounts of inhibin A and B that are similar to that measured in follicular fluid from young women.

These endocrine changes and the decline in ovarian follicles are correlated with a decrease in ovarian volume. A measurement of volume by ultrasonography that is equal to or less than 3 cm³ predicts a poor response to ovulation induction.

Testing the Ovarian Reserve

The measurement of baseline hormonal levels provides important and helpful prognostic information. Because FSH levels increase with age, elevated FSH levels on cycle day 3 (greater than 12 IU/L, but especially greater than 20 IU/L) are associated with poor performance with in vitro fertilization. A cycle day 3 FSH level that is 25 IU/L or more, or an age of 44 years or more, are both independently associated with a chance of pregnancy close to zero during ovulation induction or with assisted reproductive technologies. Values can vary between laboratories, significantly due to differences in antibodies and standards used to measure gonadotropin levels. With the most recent assays, day 3 values above 10 IU/L are abnormal and are comparable with values greater than 20 IU/L with older assays.

Keep in mind that there is no abrupt change at 40, and, therefore, these changes can apply to younger women. The age of change is determined by the rate of oocyte loss, a rate that is genetically programmed in most women. An elevated FSH at any age is correlated with fewer available oocytes, and advancing age adds the factor of lesser oocyte quality. An elevated estradiol level on day 3 (greater than 80 pg/mL) is also predictive of difficulty in achieving pregnancy. Premature estradiol elevations are associated with a hurried recruitment of follicles in response to increased FSH secretion, changes characteristic of the years preceding menopause. When both FSH and estradiol are elevated on day 3, ovarian response to stimulation is very poor.

The Clomiphene Challenge Test

Navot and colleagues developed the clomiphene challenge test as a biosassay of FSH response (which in turn reflects ovarian follicular inhibin capability). Clomiphene is administered in a dose of 100 mg/day on days 5–9. The level of FSH on day 10 is compared with the baseline level on day 3. The GnRH response of 25 IU/L or more was two standard deviations above the control values in this study, and this increase in FSH was associated with a significant prospect for failure to achieve pregnancy. Remember that there is a wide variety of FSH assays (the newer assays provide numbers that are approximately 70% of the older results), and clinicians must become familiar with the values provided by their own laboratories. There is a high incidence of abnormal responses in women over age 35, and 85% of women with increased FSH levels respond poorly to ovarian stimulation. In a general infertility population, about 10% of women have an abnormal response to the clomiphene challenge test. As expected, the percentage increased with increasing age (from 3% under age 30 to 26% over age 39), and women with unexplained infertility had a 38% incidence of abnormal tests. This test of ovarian reserve is more sensitive than just measuring basal FSH levels in predicting the potential for pregnancy in both assisted and natural reproduction. An abnormal response to the clomiphene challenge test indicates a poor prognosis no matter the woman's age. However, because age alone is an important factor (presumably due to other age-related oocyte conditions of quality), older women with normal hormone levels and normal responses to clomiphene still have diminished chances of pregnancy.

Regardless of age, women with elevated day 3 FSH levels and/or abnormal responses to the clomiphene challenge test should consider in vitro fertilization with donor oocytes. Patients should also be informed that a failure to respond to gonadotropin treatment despite normal screening hormonal levels places them at risk for progressing to ovarian failure in a relatively short time.

GnRH Agonist and Gonadotropin Combined Treatment

Recognizing that women with significant estrogen, androgen, and gonadotropin levels do not respond well to induction of ovulation, attention was turned to a method that could turn off a woman's endogenous reproductive hormone production. The availability of GnRH agonists provided such a method. The thesis underlying the use of a GnRH agonist as an adjunctive therapy in ovulation induction is straightforward: convert normogonadotropic anovulators to a hypogonadotropic hypogonad state by the process of pituitary GnRH receptor desensitization and down-regulation. Premature LH effects on the follicle and the burden of excess local androgen can be diminished and an improved therapeutic response achieved. There is reason to believe that women with anovulatory polycystic ovaries have a higher incidence of spontaneous abortions as a consequence of following induction of ovulation with gonadotropin. Combining the GnRH agonist with gonadotropin not only yields a greater pregnancy rate, but also reduces the miscarriage rate. Furthermore, premature LH release is believed to contribute to the risk of ovarian hyperstimulation in women with polycystic ovaries. A GnRH agonist with gonadotropin treatment could reduce (although not eliminate) the risk of this serious complication.

Leuprolide acetate (Lupron) is administered twice daily (0.5 mg subcutaneously) for 2 weeks. Suppression of gonadotropin secretion is confirmed by measurement of the estradiol level; a concentration less than 25 pg/mL should be achieved before treatment is initiated with gonadotropins. Lupron treatment is maintained throughout the gonadotropin regimens until HCG is administered. It is not unusual to require higher doses of gonadotropin. With this combination, no difference has been observed in the number of follicles recruited, the rate of rise and final estradiol level achieved, or the number of cycles canceled compared with gonadotropin alone. Hyperstimulation is not avoided, but per cycle fecundity appears to be modestly increased. Any of the available GnRH agonists can be used for this purpose. Pituitary suppression can be combined with highly purified or recombinant FSH; it appears that very little LH is necessary for normal responses.

The administration of a GnRH agonist to a woman who has menstrual function will initially produce a stimulatory response, known as the "flare." The magnitude of the flare response depends upon when in the cycle the agonist is administered. During the follicular phase or in anovulatory women, the flare is greater, and enlarged follicular cysts can occur. This response can be minimized by beginning therapy during the midluteal phase or by administering a progestational agent (e.g., 10 mg medroxyprogesterone acetate daily for 10 days) and beginning GnRH agonist treatment after 3 days of progestin.

During the hypostrogenic period of time, menopausal-like symptoms are common, especially hot flushes. Forewarning and reassurance are usually sufficient to help patients tolerate these short-lived reactions.

Utilization of a GnRH agonist suppresses endogenous LH levels to such a low level that, after ovulation, the corpus luteum requires additional exogenous support. One can administer HCG (2000 IU) twice, 3 days, and 6 days after ovulation, or progesterone supplementation beginning 3 days after ovulation (4 days after the LH surge): intravaginal progesterone suppositories, 25 mg twice per day, intramuscular progesterone, 50 mg per day, or oral micronized progesterone, 300 mg per day. The dose of micronized progesterone is relatively high and should be administered at bedtime to avoid side effects. Progesterone in a gel that adheres to the vaginal mucosa (Crinone) is available in a prefilled vaginal applicator, with one 8% (90 mg progesterone) application daily. Vaginal administration accomplishes targeted delivery to the uterus without producing high circulating levels. The pregnancy rates with either HCG or progesterone treatment are the same, but the use of HCG adds to the risk of hyperstimulation.

GnRH Antagonists

More extensive alterations in the GnRH structure yield compounds that bind to the GnRH receptor and totally block any stimulation of the pituitary gonadotropin-secreting cells. GnRH antagonists can be used concomitantly with gonadotropin induction of ovulation to prevent premature LH surges. Because GnRH antagonists are competitive inhibitors, a GnRH agonist could be used for the ovulatory signal, possible avoiding the use of HCG and reducing the risk of ovarian hyperstimulation. The optimal use of GnRH antagonists is currently being explored.

Adding Growth Hormone

As discussed in Chapter 6, the primate ovarian follicular cycle is very dependent upon autocrine/paracrine factors. In view of the critical role played by insulin-like growth factor-I, it was quickly recognized that the addition of growth hormone (which stimulates insulin-like growth factor-I production), usually in a dose of 24 IU given...
Induction of Ovulation With Gonadotropin-Releasing Hormone (GnRH)

Numerous advantages can be cited in favor of pulsatile GnRH therapy for ovulation induction. GnRH methodology, once established, is simple to use, requires no extensive (or expensive) follicular monitoring, and, relative to its counterparts, is quite safe. Ovarian hyperstimulation and multiple gestations are rare because only "physiologic" levels of FSH should be generated. Because GnRH serves largely a permissive role, the internal feedback mechanisms between the ovary and pituitary should be operative, yielding follicular growth and development similar to a normal menstrual cycle in response to the "turning on" of the system by GnRH.

GnRH is administered constantly in a pulsatile fashion by a programmable portable minipump. Induction of ovulation with the GnRH pump is most effective in women with hyperthalamic amenorrhea (absence of menstrual bleeding following a progestin challenge) where endogenous GnRH is dysfunctional or absent. Unfortunately, although it has been successful in some anovulatory women with polycystic ovaries, this success is far less than anticipated, and these patients are at greater risk for hyperstimulation and multiple pregnancy, therefore requiring a lower dose (2.5 µg per bolus). Nevertheless, the GnRH pump is a safer, less expensive alternative for these patients than gonadotropin therapy. The GnRH pump is also effective in women with hyperprolactinemia, providing a good alternative if dopamine agonist treatment cannot be tolerated.

Ovarian Surgical Procedures

Irving Stein and Michael Leventhal, in 1935, at the Michael Reese Hospital in Chicago, described 7 cases of the syndrome that bears their names. Ovarian Surgical Procedures

Chapter 6

Adding Growth Hormone

As discussed in Chapter 6, the primate ovarian follicular cycle is very dependent upon autocrine/paracrine factors. In view of the critical role played by insulin-like growth factor-I, it was quickly recognized that the addition of growth hormone (which stimulates insulin-like growth factor-I production), usually in a dose of 24 IU given intramuscularly every other day, might facilitate ovulation induction by gonadotropins, and especially, might convert poor responders to good responders. Initial results in poor responders were favorable, although this may be limited to patients with polycystic ovaries. In women with normal responses, however, the addition of growth hormone did not improve results or reduce the total dose of gonadotropin. Thus, in normally responding women, the growth hormone and IGF-I system may be operating at maximal levels. In poorly responding women, concomitant treatment with growth hormone can reduce the amount of gonadotropin necessary, shorten the duration of treatment, and increase the chance of pregnancy. However, not all poorly responding women respond favorably to growth hormone treatment. The effective selection of patients, dosage, and treatment regimens remains to be standardized, and the extreme cost of growth hormone is a significant disadvantage. Another approach is to administer growth hormone-releasing hormone. In a small randomized trial (13 women), the addition of growth hormone-releasing hormone to gonadotropin treatment was associated with a higher pregnancy rate. However, others have administered 500 µg of growth hormone releasing hormone twice daily to poor responders with little effect, although the dosage of gonadotropin may be reduced.
Today, a new type of "wedge resection" is available. Using either cautery, diathermy, or laser vaporization by means of the laparoscope, destruction of ovarian tissue at multiple sites (15–20 per ovary) can achieve spontaneous ovulations or an increased sensitivity to clomiphene, ultimately achieving pregnancy rates equivalent to those with gonadotropin treatment (but without the risks of hyperstimulation and multiple pregnancies). These procedures are associated with the same decrease in androgens and inhibit as observed with wedge resection. When clomiphene is reinstituted, 70–80% of patients will ovulate, and approximately 60% will achieve pregnancy. Adhesion formation remains a problem, but perhaps it is less profound than in the traditional surgery. Second look laparoscopy with lysis of adhesions is indicated if pregnancy does not follow successful ovulations. This therapy can be performed at the time of a planned diagnostic laparoscopy in a patient with known anovulation and polycystic ovaries. These procedures are worth considering by patients who are reluctant to pursue the more expensive and difficult methods of the assisted reproductive technologies. These methods are the modern equivalent of the original ovarian wedge resection pioneered by Stein and Leventhal.

Superovery and IUI for Unexplained Infertility

The methods of assisted reproductive technology and superoverification with intrauterine insemination (IUI) increase the prospect of pregnancy in couples with unexplained infertility. The results with superoverification alone are inferior to those achieved with one of the assisted reproductive techniques, but the overall cost is also less. Studies indicate that subfertile abnormalities do exist in women with unexplained infertility. These changes (elevated gonadotropins and estradiol in the follicular phase and decreased luteal phase progesterone) have been attributed to a reduction in ovarian reserve as usually seen in older women. Thus, there is a rational basis to provide ovarian stimulation.

A cumulative pregnancy rate of 40% can be achieved in couples with unexplained infertility after 6 cycles of superoverification or 3 cycles of in vitro fertilization. In randomized, controlled clinical trials, the monthly pregnancy rates in couples with unexplained infertility is increased 2–3-fold (a monthly fecundity rate of 9%) with clomiphene treatment, and with human gonadotropins, the monthly fecundity rate is approximately 10–15%. Although some studies have failed to yield increased pregnancy rates with the empiric use of clomiphene, we believe couples with unexplained infertility should be offered superoverification or one of the assisted reproductive technologies. However, superoverification should be limited to 3 or 4 cycles.

The addition of intrauterine insemination of washed sperm (IUI) to clomiphene treatment increased fecundity in couples with unexplained infertility or surgically treated endometriosis, compared with intercourse without hormonal stimulation. As with almost all infertility, success is age-related, with pregnancy rates 3–4-fold greater in women under age 20. Although the combination of IUI with clomiphene may enhance the pregnancy rate in couples with unexplained infertility, IUI by itself does not.

The combination of IUI with gonadotropin stimulation provides even better results than clomiphene and IUI. In two studies, the pregnancy rate per cycle with gonadotropins and IUI was 19% and 14.3% compared to 4% and 7.7% with clomiphene and IUI respectively. In a meta-analysis of randomized trials comparing gonadotropin stimulation with IUI or with timed intercourse, IUI was associated with approximately a 2-fold greater pregnancy rate. In a randomized trial, superoverification and IUI resulted in a 3-fold greater pregnancy rate compared with intrauterine insemination and 2-fold greater compared with either IUI alone or a combination of superoverification and intracervical insemination.

The effectiveness of gonadotropin treatment combined with IUI has been repeatedly demonstrated for a variety of infertility disorders, including endometriosis, cervical factor, and unexplained infertility. As with clomiphene, few pregnancies occur after 4 cycles of treatment.

In summary, a meta-analysis of methods used to treat unexplained infertility, pregnancy rates were estimated to be as follows:

<table>
<thead>
<tr>
<th>Pregnancy Rate per Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>No treatment</td>
</tr>
<tr>
<td>IUI</td>
</tr>
<tr>
<td>Clomiphene</td>
</tr>
<tr>
<td>Clomiphene &amp; IUI</td>
</tr>
<tr>
<td>Gonadotropins</td>
</tr>
<tr>
<td>Gonadotropins &amp; IUI</td>
</tr>
<tr>
<td>IVF</td>
</tr>
</tbody>
</table>

Because it is relatively inexpensive, easy to administer, and not usually burdened with major side effects (except for a 2-fold increase in twinning), clomiphene is usually the first treatment choice. Patient demand for this treatment is substantial, and it is unreasonable to respond to that demand with 3–4 clomiphene cycles. The uncertainty regarding the efficacy of adding IUI to clomiphene influences us to proceed from clomiphene treatment to the combination of gonadotropins and IUI, with an expectation of achieving a pregnancy rate of 15% per cycle. The combination of clomiphene and IUI is still appropriate in circumstances where cost, time, and avoidance of complications are paramount. The use of IUI alone or gonadotropins without IUI is not recommended for the treatment of unexplained infertility.

The effects of age remain critical whatever the method of ovarian stimulation. In one series of 136 treatment cycles, no pregnancies were achieved in women who were 43 or older. Despite rare successes in older women, it borders on the exploitive to treat women over age 40 with superoverification and IUI. Indeed, pregnancy rates begin to decline with this treatment around age 35.

The incidence of multiple births and ovarian hyperstimulation is higher with this empiric treatment. Patients and clinicians undertaking empiric superoverification should have ready recourse to in vitro fertilization. If gonadotropin stimulation is excessive, several choices are available: avoiding HCG administration and cancelling the cycle, aspirating most of the ovarian follicles with ultrasound guidance and proceeding with in vitro fertilization.

Fertility Drugs and the Risk of Ovarian Cancer

An analysis of case-control studies of ovarian cancer conducted in the U.S. from 1987–1989 concluded that the risk was increased among women who had used fertility drugs and among women who had been exposed to long periods of sexual activity without contraception. However, this conclusion was limited by the small number of cases, and the exact drugs and dosage used by these cases were unknown. Concern over this issue was further fueled by a report that prolonged use of clomiphene (12 or more cycles) was associated with an increased risk of borderline or invasive ovarian tumors. The conclusion in this comparison of a cohort of women with the rate of ovarian tumors in the general population was based on only 5 cases of borderline tumors and 4 cases of invasive tumors. Nevertheless, the idea that superovulation can increase the risk of ovarian cancer is biologically plausible because there is a decreasing risk for invasive epithelial ovarian cancer associated with those conditions marked by a decrease in the number of ovolutions: multiparity, increasing duration of breastfeeding, and the use of oral contraception.

It is worth noting that a long-term follow-up of 2632 women in Israel who had received drugs for induction of ovulation did not record an increased incidence of ovarian cancer in the subsequent years. More recently, stimulated by the above reports, several epidemiologic studies appeared with the strength of larger case numbers. A study from Israel indicated an increased risk, but the relative risk was not statistically significant. Two case-control studies from Italy and a Danish case-control study (the largest in size) failed to find a relationship between fertility drugs and the risk of epithelial ovarian cancer.

The best evidence does not indicate an increased risk of ovarian cancer with less than 12 months of clomiphene treatment (there is no reason to pursue clomiphene treatment for more than 12 months), or with gonadotropin treatment. Nulliparity and infertility in nulliparous women are associated with a slight increase in the risk of ovarian cancer, but ovarian stimulation with fertility drugs does not change this baseline risk.

Chapter References

3. Ernst S, Hite G, Cantrell JS, Richardson Jr A, Benson HD. Stereochiometry of geometric isomers of clomiphene: a correction of the literature and a reappraisal of structure-activity present). It should truly be a rare patient in whom wedge resection of the ovaries is necessary.
Treatment of prolactin-secreting macroadenomas with the once-weekly
Improvement of spontaneous pregnancy rate after stopping gonadotropin therapy for anovulatory infertility,
A comparative prospective study of conventional regimen with chronic low-dose administration of follicle-stimulating hormone for anovulation associated
March CM, Tredway DR, Mishell Jr DR,
Value of bromocriptine in unexplained primary infertility: a double-blind controlled trial,
DeVane GW, Guzick DS,
Fernandez H, Coste J, Job-Spira N,
Aboulghar MA, Mansour RT, Serour GI, Rizk P, Riad R,
Hypersecretion of luteinizing hormone and ovarian steroids in women with recurrent
Association of moderate obesity with a poor pregnancy outcome in women with polycystic ovary syndrome treated with low
Kol S, Levron J, Lewit N, Drugan A, Itskovitz-Eldor J,
The use of vaginal ultrasound for monitoring endometrial preparation in a donor oocyte program,
Ovarian hyperstimulation syndrome: report of a case with notes on pathogenesis and treatment,
Akin JW, Shepard MK,
Ho ML, Chen JY, Ling UP, Chen JH, Huang CM, Chang CC, Su PH,
Predictive value of serum oestradiol concentrations and oocyte number in severe ovarian hyperstimulation syndrome,
Vascular endothelial growth factor levels in serum and follicular fluid of patients undergoing in vitro fertilization,
Dickey RP, Olar TT, Taylor SN, Curole DN, Matulich EM,
Spontaneous pregnancy and its outcome after human menopausal gonadotropin/human chorionic
treatment,
Ritchie WGM,
Use of gonadotropins in ovulation induction,
Haning Jr RV, Boehnlein LM, Carlson IH, Kuzma DL, Zweibel WJ,
Engel T, Jewelewicz R, Dyrenfurth I, Speroff L,
The natural history of multiple pregnancies after assisted reproduction: is spontaneous fetal demise a clinically significant
Shoham Z, Di Carlo C, Patel A, Conway GS, Jacobs HS,
Mashiach S, Bider D, Moran O, Goldenberg M, Ben-Rafael Z,
Schenker JG,
Cabergoline. A review of its pharmacological properties and therapeutic potential in the treatment of hyperprolactinemia and inhibition of lactation,
Leerentueld R, Van Gent I, Der Stoep M, Wladimiroff J,
The impact of insulin resistance on the outcome of ovulation induction with low-dose follicle stimulating hormone in women with polycystic ovary syndrome,
Kol S, Levron J, Lewit N, Drugan A, Itskovitz-Eldor J,


Copyright ©
Assisted reproductive technology (ART) refers to all techniques involving direct retrieval of oocytes from the ovary. The first and still most common procedure is in vitro fertilization, but there is an ever increasing list of technologies.

IVF — In Vitro Fertilization: extraction of oocytes, fertilization in the laboratory, transcervical transfer of embryos into the uterus.

GIFT — Gamete Intrafallopian Transfer: the placement of oocytes and sperm into the fallopian tube.

ZIFT — Zygote Intrafallopian Transfer: the placement of fertilized oocytes into the fallopian tube.

TET — Tubal Embryo Transfer: the placement of cleaving embryos into the fallopian tube.

POST — Peritoneal Oocyte and Sperm Transfer: the placement of oocytes and sperm into the pelvic cavity.

In addition, techniques of sperm retrieval and sperm injection are now part of the assisted reproductive technology armamentarium:

ICSI — Intracytoplasmic Sperm Injection (of a single spermatozoon).

TESE — Testicular Sperm Extraction.

MESA — Microsurgical Epididymal Sperm Aspiration.

It has been over two decades since the birth of the first child conceived by in vitro fertilization (IVF). During the intervening years, the number of IVF programs has increased to over 250 in the United States alone, technology has evolved, the success rate has improved dramatically, and indications for IVF have enlarged. In addition, procedures that utilize some, but not all, of the methodology of IVF have become a part of clinical practice.

Numerous volumes have been published which extensively cover all aspects of ART, and we will not duplicate that effort. Rather, this chapter will focus on the evolving techniques with an emphasis on ICSI, and will review briefly the place of ART in the treatment of nontubal disease.

Patient Selection for IVF

The initial experience with in vitro fertilization involved women with tubal disease, but early in the 1980s, the treatment was extended to individuals with male factor infertility, unexplained infertility, endometriosis, and immunologic causes for infertility. In addition, a high degree of success has been obtained using donor oocytes for women with premature ovarian failure or decreased ovarian function. Although it is reasonable to recommend tubal surgery in young women with mild distal tubal disease, IVF is the treatment of choice for patients with more severe distal disease, proximal obstruction (especially after 6 months have elapsed following cannulation or balloon tuboplasty), and for patients who have failed to achieve pregnancy within 2 years after tubal surgery or when tubal obstruction persists after surgery. Large hydrosalpinges can reduce the success rate with IVF, and removal is recommended prior to treatment.

Couples undergoing IVF should be screened for antibodies to HIV-1/2 and HTLV-1, hepatitis B antigen, and antibodies to hepatitis B and C. Testing for chlamydia, syphilis, gonorrhea, and cytomegalovirus should also be considered. Smoking, as always, should be discouraged; smokers require twice as many IVF cycles to achieve pregnancy.

Whereas IVF can overcome a number of the barriers to fertility, it suffers the same limitation as does in vivo fertilization when it comes to age (discussed in Chapter 26). In successful IVF programs, where the delivery rate per oocyte retrieval can be over 30% in women age 35 or younger, the figure is 10% or less in women age 40 and over. Beyond the simple effect of age on oocyte quality, there also is a negative influence on IVF success from decreased ovarian responsiveness. This can be manifested by poor response to exogenous gonadotropin stimulation with abnormal hormone profiles and the retrieval of small numbers of oocytes.

There is no exact definition of a poor responder, but it encompasses those who respond to stimulation with the development of 4 or fewer follicles or with depressed estrogen levels. Pretreatment diagnosis of these individuals can be achieved by measuring basal FSH and estradiol levels, and by using the clomiphene challenge test (see Chapter 26).

Three distinct populations have been identified on the basis of the cycle day three FSH. With FSH values less than 15 IU/L, the pregnancy rate with IVF was 24%, whereas when the FSH values were 15 to 24.9 IU/L, the pregnancy rate was 13.6%. When the FSH was 25 IU/L or higher, the initial pregnancy rate was 10.7%, and the ongoing pregnancy rate was only 3.6%. Remember in all discussions pertaining to specific hormone levels that these may differ between laboratories, depending on which assay system is used. Assays using new standards have an upper limit of normal of 10–12 IU/L compared with the 25 IU/L cited in the above study.

Older women (certainly 40 and older, but perhaps even women in their late 30s) with infertility and younger women with abnormal basal FSH and estradiol values and an abnormal response to a clomiphene challenge should seriously consider an early resort to hormonal stimulation for both multiple oocyte response and better support of the endometrium. Older couples should be provided the option of oocyte donation from young donors instead of standard assisted reproductive technologies.

Stimulation Protocol

The momentous work of Edwards and Steptoe that produced the first birth from IVF was developed over more than 10 years. Success was finally achieved by utilizing a nonstimulated cycle with timing of oocyte retrieval based on measurements of luteinizing hormone (LH) at 3-hour intervals. Nonstimulated cycles are still used as a means of decreasing expenses, but the delivery rate per retrieval is only approximately 6%. The very low success rate associated with nonstimulated cycles led to the use of clomiphene citrate and human menopausal gonadotropins to stimulate the development of multiple ovarian follicles. Injections of human chorionic gonadotropin (HCG), whose biologic activity mimics that of LH, were utilized to allow more certain timing of oocyte retrieval.
In the late 1980’s, gonadotropin-releasing hormone agonists (GnRH agonists) were introduced as a means of down-regulating the pituitary to prevent premature ovulation, which in the past had necessitated canceling approximately 15% of IVF cycles prior to egg retrieval. The down-regulating effects of GnRH agonists, as opposed to the stimulatory effects of GnRH, are related to the frequency of administration and the prolonged occupation of GnRH receptors by the agonists. Since their introduction, pregnancy rates have increased because of the opportunity to retrieve more eggs and because of the increase in the number of oocytes obtained in GnRH agonist cycles. However, use of an agonist increases the amount of gonadotropins needed to stimulate follicular growth, and, thus, it also increases the expense. Despite this negative aspect, the combination of GnRH agonist and gonadotropin controlled hyperstimulation of the ovary, and substitution of exogenous HCG for the endogenous LH surge, is now utilized by most IVF programs. This approach has another beneficial attribute. Because the fear of premature ovulation is virtually eliminated, there is more flexibility in scheduling the necessary interventions.

Protocols starting standard doses of GnRH agonists in the follicular phase of the cycle preceding the one in which gonadotropin is started have the drawback, because of an initial stimulatory effect, the flare, of the GnRH agonist, of causing a larger number of ovarian cysts and higher androgen levels than cycles in which the agonist is started in the midluteal phase of the cycle. In a further variation, GnRH agonist is started on the second or third day of menses, and gonadotropin is started 2 days later. This “flare” protocol tries to capture the initial stimulatory effect of the GnRH agonist and build on it with gonadotropin before the inhibitory effect of the agonist is established. This flare approach is particularly useful for poor responder patients. To aid timing, a common pretreatment for women with irregular menses uses oral contraceptives for one or two cycles.

Most commonly, the GnRH agonist is started in the midluteal phase, and gonadotropin treatment is started following menses. Ultrasonography is used prior to initiation of gonadotropin treatment to rule out the presence of an ovarian cyst larger than 15 mm. Ovarian cysts form in approximately 10% of women when the GnRH agonist is started in the midluteal phase, but these cysts almost always regress spontaneously in 1–3 weeks. Gonadotropin treatment is postponed until the cysts disappear or decrease to less than 15 mm in size.

GnRH agonist administration is continued for the duration of gonadotropin treatment. Whereas leuprolide (Lupron) injected subcutaneously is the most commonly used GnRH agonist in the United States, equally good results, if not better, have been achieved with the use of nafarelin (Synarel), which is administered as a nasal spray, one inhalation bid. The initial dose of gonadotropin is usually 225–300 IU/daily, except in young women or those with polycystic ovarian disease where a lower dose (75–150 IU/daily) is appropriate. The dose can be adjusted as cycle monitoring proceeds with ultrasonography and estradiol measurements.

Intramuscular injections of urine gonadotropins containing both FSH and LH were the mainstays of treatment until the development of a urinary gonadotropin that contained primarily FSH (see Chapter 30). One of the goals in limiting or eliminating LH is to decrease stimulation of the androgenic component of the ovarian response because excessive androgens block ovarian follicular growth and development. However, in clinical practice, there was no clear-cut advantage for the reduced LH product. Still, recent advances in the advanced urinary gonadotropin technology have made the reduced LH product a distinct option.

Recombinant FSH also allows subcutaneous administration. Differences comparing recombinant FSH with the urinary hormones include enhanced purity, the need for less medication, the achievement of slightly lower estradiol levels, and a modest increase in the pregnancy rate.

Whenever product or protocol is used, there is a 10–15% cancellation rate because of inadequate follicular response. Measurements of basal FSH and estradiol and the clomiphene challenge test can provide information for discussing with patients the expectations for success. Ideally, a stimulation protocol can be tailored to boost the chances for an adequate response. However, it is evident that most poor responders are resistant to conversion to good responders. The 5 most common changes in protocol designed to accomplish that elusive goal are:

1. Increase the gonadotropin dose or change the mix of FSH-LH and FSH. This is usually not successful. No improvement was noted when the gonadotropin dose was increased from 300 to 450 IU beginning on day 1 or 2 of the cycle.
2. Use of the flare protocol, including the use of a microdose flare (20 µg leuprolide bid), approximately 1/50 the usual dose.
3. Lower the dose of the GnRH agonist to allow some pituitary action, but starting as usual in the midluteal phase.
4. Omit the GnRH agonist and use gonadotropin alone or a combination of clomiphene and gonadotropin. To prevent premature ovulation in the absence of down-regulation by the GnRH agonist, a single dose of a GnRH antagonist can be given at midcycle to prevent the LH surge.
5. A less common approach, limited by expense and availability of drug, is the use of growth hormone in conjunction with gonadotropin. Moreover, results with this combination have been mixed; a double-blind study found no benefit associated with the addition of growth hormone.

Monitoring Ovarian Response

Measurements of serum estradiol and ultrasonography of ovarian follicles are used to monitor the ovarian response to stimulation. The minimum goal of stimulation is to achieve the growth of a lead follicle to at least 18 mm diameter, and to have at least 3 or 4 other follicles with diameters of 14 mm or greater, combined with estradiol levels of approximately 200 pg/mL per large (14 mm or greater) follicle. Once this level of stimulation is achieved, a single injection of 5000 or 10,000 IU of HCG is given to induce final follicular maturation. The time interval between HCG injection and retrieval is critical. Whereas 34–36 hours is standard and believed to allow complete follicular growth, most protocols now allow for at least 48 hours, thereby increasing the risk of hyperstimulation. The values given for follicle size and estradiol levels on the day of decision for HCG injection are only rough guidelines because ultrasound measurements can differ among observers and machines. In addition, each program must establish, based on its own experience, its criteria for determining the adequacy of follicle size. Many programs allow slight variations from one laboratory to another, and, therefore, are not compatible with universalization. Estradiol levels per follicle can vary widely and still be compatible with successful IVF. Concern has been expressed that elevated levels of progesterone, signifying premature luteinization, just prior to HCG injection can disrupt the cycle. The weight of evidence indicates that the frequently seen modest rise in progesterone does not interfere with pregnancy rates.

To avoid the significant medical risks of hyperstimulation, cancellation of the cycle and avoidance of HCG injection should be strongly considered if the ovaries are markedly hyperstimulated (more than 25 follicles or a total estradiol greater than 4000 pg/mL). (Hyperstimulation is discussed in Chapter 30.) Again, the specific values may not be valid with other assays, and each program must establish which estradiol level is its warning sign. Alternatives to cancellation include proceeding with the regular IVF cycle in the face of markedly elevated estradiol levels, but cryopreserving all the embryos for transfer in subsequent cycles to avoid further HCG stimulation accompanying a pregnancy, or discontinuing the daily gonadotropin injections (coasting) and delaying retrieval until the estradiol level (measured daily) drops to 2500–3000 pg/mL or lower, at which point HCG is administered. However, cycles in which coasting produces very precocious drops in estradiol to below 1000 pg/mL warrant cancellation. The pregnancy rates with coasting are lower than those obtained with the usual methods.

The risk of hyperstimulation can be decreased by lowering the dose of gonadotropin used to initiate the cycle. The use of FSH as opposed to mixtures of FSH and LH in women with polycystic ovaries is not protective; with all stimulation protocols, the key is careful use of appropriate doses. An important protection against hyperstimulation is the use progesterone after retrieval, rather than HCG injections, to support the luteal phase. Retrieval itself, with aspiration of follicular fluid and granulosa cells, is somewhat, but not absolutely, protective against hyperstimulation.

When follicles are fully developed, the endometrium viewed by ultrasonography should be at least 8 mm wide. If the width is 6 mm or less, there is a reduced chance of successful IVF. Concern has been expressed that elevated levels of progesterone, signifying premature luteinization, just prior to HCG injection can decrease or decrease to less than 15 mm in size.

Oocyte Retrieval

Ultrasonically guided vaginal oocyte retrieval is performed approximately 34–35 hours after the HCG injection, but as late as 39 hours with the combined use of a GnRH agonist and gonadotropin. The HCG injection allows for confidence in this precise timing. Intravenous analgesia and/or light anesthesia provide an acceptable level of comfort. In all cases, monitoring by a pulse oximeter is mandatory, as is immediate access to emergency resuscitation equipment. A number 16 or 17 long needle is placed down a sterile needle guide that is attached to the upper side of the vaginal ultrasound transducer. A line on the monitor screen indicates the path the needle will traverse once it enters the peritoneal cavity and the ovary. The ultrasound transducer is manipulated to position a follicle along this pathway. Usually only one puncture of each ovary is needed to allow sequential aspiration of the follicles. Accessibility of the ovaries is rarely a problem. Rare complications of the procedure are intra-abdominal bleeding or introduction of infection into the ovary and pelvis.

Oocyte Culture

Ultrasonically guided vaginal oocyte retrieval is performed approximately 34–35 hours after the HCG injection, but as late as 39 hours with the combined use of a GnRH agonist and gonadotropin. The HCG injection allows for confidence in this precise timing. Intravenous analgesia and/or light anesthesia provide an acceptable level of comfort. In all cases, monitoring by a pulse oximeter is mandatory, as is immediate access to emergency resuscitation equipment. A number 16 or 17 long needle is placed down a sterile needle guide that is attached to the upper side of the vaginal ultrasound transducer. A line on the monitor screen indicates the path the needle will traverse once it enters the peritoneal cavity and the ovary. The ultrasound transducer is manipulated to position a follicle along this pathway. Usually only one puncture of each ovary is needed to allow sequential aspiration of the follicles. Accessibility of the ovaries is rarely a problem. Rare complications of the procedure are intra-abdominal bleeding or introduction of infection into the ovary and pelvis.
Oocytes, surrounded by their cumulus masses, are identified under either a dissecting or an invert microscope. Exposure to ambient temperature and room air is minimized. One of the major breakthroughs in IVF was the discovery that sperm need not be added to the eggs immediately after retrieval; oocytes have a higher chance for fertilization if insemination follows retrieval by 4 to 6 hours. 2

Sperm are prepared by washing, centrifugation, discarding the supernatant, and overlaying the sperm pellet with fresh media. Sperm that swim up into the media are used for insemination, with approximately 50,000 sperm (100,000 or more if the sperm specimen is poor) placed in each petri dish containing an oocyte. Isolation of the most motile sperm can also be accomplished by separating the sperm on columns of liquid albumin, hyaluronate, glass wool, or gradients of silica particles. 2

A variety of media has been used for embryo culture. Most media for IVF contain a protein source, such as maternal serum, fetal cord serum, albumin and/or serum globulin. All protein components require screening for HIV and hepatitis. Because proteins, on occasion, may be deleterious to embryos, protein free media have been successfully used in limited studies. 2

Culture of oocytes and embryos on cell monolayers from a variety of sources has been advocated as a means of enhancing development. 2 Pregnancy rates were increased and blastomere fragmentation reduced when human embryos were cultured on outgrowths of bovine tissue. 2 The explanation for the beneficial effect could be the secretion of an “embryotrophic” factor by the somatic cells or their removal of detrimental components from the media. Bavister suggested that greater attention to culture media quality can yield the same gains as those seen with co-culture, and this has been confirmed by recent experience. 2 In addition, co-culture with granulosa cells has not helped couples with previous fertilization failures. 2

Fertilization

The day after insemination, cumulus cells that remain attached to the zona pellucida are removed, and the egg is examined for evidence of fertilization (the presence of 2 pronuclei). 65–80% of mature oocytes will fertilize. Approximately 6% of oocytes contain 3 or more pronuclei, indicating penetration by more than one sperm. It is important to identify and discard these abnormal embryos at this early stage because they have the potential to cleave and appear normal at later preimplantation stages. If the diagnosis is missed at the pronuclear stage, the embryos will be transferred and the result will be a failure to implant or a spontaneous miscarriage.

A failure to fertilize is sometimes viewed by clinicians as a “human egg test” for that specific couple. However, fertilization failure in one cycle, despite normal sperm and eggs, is not an absolute indication of an insurmountable problem. Indeed, a high rate of fertilization has been reported in subsequent cycles after an initial fertilization failure, although the eventual pregnancy rate is a little lower than the usual rate in IVF programs. 2 To demonstrate a complete failure of fertilization, at least 3 IVF cycles are necessary. Surprisingly the use of ICSI in a cycle following one with unexplained fertilization failure does not always overcome the failure to fertilize.

Extra embryos can be cryopreserved at the pronuclear stage. There is no known limit on duration of cryopreserved embryo storage. About two-thirds of embryos survive the freezing and the subsequent thawing process. The transfer of cryopreserved embryos adds significantly (15% delivery per transfer) to the success rate with a single retrieval and lowers the cost. Older age of the woman is associated with a decrease in pregnancy rates, not only because of a decrease in the number of oocytes and a reduction in oocyte and embryo quality, but also because fewer embryos are available for cryopreservation. 2

Embryo Transfer

Embryos have been transferred successfully at any stage from the pronuclear to the blastocyst, although most commonly, they are transferred when development is between the 8- and 10-cell stage, approximately 72–80 hours after retrieval. Transfer of more than one embryo increases the chances for pregnancy, but the multiple pregnancy rate with IVF is approximately 35%. 2

High order multiples can create unfortunate outcomes for the pregnancy, the children, and for subsequent family life. It is not unusual for patients to welcome the prospects for a multiple birth, but they must be informed of the associated risks. To decrease the incidence of multiple birth, programs in the U.K. transfer a maximum of 2 embryos. The guidelines of the American Society for Reproductive Medicine are more liberal, and they are influenced by the age of the woman. The recommendation is for no more than 3 embryos to be transferred in women younger than age 35, no more than 4 at age 35–40, and no more than 5 in women older than age 40. Our preference is to transfer 2 good quality 8–10 cell embryos in women under the age of 35, whereas between age 35–39 may have 2–3 good quality embryos transferred. It may be reasonable to place higher numbers in women age 40 and older, but this is influenced by the experience of the individual program. At least one study did not experience a decrease in multiple birth risk in older women. 2 While we cite “good quality” of the embryos as a qualification, it must be recognized that this assessment is highly subjective. On occasion, transfer of “poor” quality embryos results in pregnancy. The ability to use fetal reduction to limit the number of continuing fetuses lessens somewhat the concerns over the major hazards of multiple pregnancies. In experienced hands, the risk of losing the entire pregnancy from selective fetal reduction is less than 10%. 2 However, moral concerns and the psychological impact of fetal reduction are important considerations.

Success with culture to the blastocyst stage has raised the hope that high pregnancy rates can be achieved with the transfer of only one blastocyst (reaching this stage in culture is predictive of excellent potential). 2

Ancillary medications given around the time of transfer, such as prostaglandin inhibitors, tranquilizers, or antibiotics are of no proven value. It is common practice to supplement the luteal phase with progesterone. Several options are available, including progesterone vaginal suppositories, 50–100 mg daily, progesterone in oil, 50–100 mg im daily, one application of vaginal progesterone daily, or oral micronized progesterone, 200 mg tid. An alternative therapy is the use of HCG, 1500 IU every 3 days for 4 doses; however, HCG supplementation increases the risk of ovarian hyperstimulation. The value of luteal phase supplementation can be questioned, but there is no doubt that it provides some psychologic benefit for both patient and physician. An initial quantitative HCG measurement can be obtained 12 days after transfer. If the test is positive, a follow-up test is obtained 3 days later. This will establish the trend of the HCG and provide prognostic information.

A rise in the HCG level indicates pregnancy, whereas a drop indicates a failed cycle. If weekly determinations demonstrate appropriate elevations, ultrasonography 5 weeks after the transfer should reveal a fetal heart beat. If not, repeat ultrasonography can be performed in one week. The presence of a rapid heart beat is an excellent prognostic sign, and if the woman is under age 30, the chance for a spontaneous miscarriage is near zero; between age 31–35, the risk of spontaneous miscarriage is 4%. 2 If the woman is over age 40, the prognosis is less sanguine, with spontaneous miscarriage rates close to 20%. If the woman has a history of recurrent early pregnancy losses, a positive heartbeat still leaves the chance of a spontaneous miscarriage in the 20–25% range.

Assisted hatching consists of making an opening in the zona pellucida just prior to transfer to help the embryo emerge. This can be accomplished by sitting open the zona with a microneedle, by treating a localized area of the zona with an acid Tyrode’s solution, or by using a laser to breach the zona. Assisted hatching has been reported to increase the implantation rate in poor prognosis patients, especially in older women. 2 However, two studies, one in women older than age 36, and the other in women older than age 38, could find no benefit with assisted hatching. 2

IVF Results

Some skepticism has been expressed concerning the pregnancy rates reported by IVF programs. 2 Competition for patients is a motivation for some of the exaggerations and half-truths contained in advertisements and public statements. An associated confusion is generated by the differing criteria that can be used to determine what constitutes a success. A slight unsustained rise in the HCG is properly termed a “chemical pregnancy,” and it should not be counted as a success. The most important indicators of success are the delivery rate per retrieval and the delivery rate per cycle initiated. Even these figures cannot be properly evaluated without knowledge of the specific age mix of the treated women. Obviously, a program with a higher percentage of older women should not be expected to do as well as one with mostly younger women. In addition, clinics that transfer large numbers of embryos to each woman can marginally inflate their pregnancy numbers but at a significant price of multiple pregnancies.

Uniform reporting by clinics, monitored by the Society for Assisted Reproductive Technology, has produced yearly statistics available to the public. 2 Results from the U.S. and Canada are published nearly 3 years later, reflecting a significant lag because of the need to await deliveries resulting from cycles initiated and the
In 1995, the deliveries per retrieval were 22.3% for standard IVF, 27.0% with GIFT, and 27.9% with ZIFT. Women under age 35 had a 27.2% delivery rate per retrieval with IVF compared to 10.5% in women older than age 39. With transfers of donor embryos, the delivery rate was 36.0%, whereas with the transfer of cryopreserved embryos, the delivery rate was 15.2%. These figures for 1995, reported in 1996, are probably already outdated. Many clinics are achieving delivery rates in the 35% range, especially in women under age 35. The advantages for GIFT and ZIFT disappear when comparisons are made within the same program.

Surprisingly, 3% of pregnancies achieved through IVF are ectopic, emphasizing the need for close ultrasonographic and HCG titer surveillance (see Chapter 32). There is uncertainty regarding the role of Y chromosome microdeletions, because of the incidence of combined pregnancy among patients who have undergone one of the assisted reproduction procedures is much higher, closer to 1 in 100 pregnancies. A case-control study has concluded that the risk of ectopic pregnancy with assisted reproduction is due to the multiple oolvations and high hormone levels secondary to the stimulation protocols.

Twenty percent of clinical pregnancies result in spontaneous miscarriages, which is close to the rate in infertility populations; there is no increase in the rate of congenital malformations. The children experience normal growth and development.

The multiple pregnancy rate is approximately 35% (30% twins, 5% triplets, and about 0.6% higher multiples). This leads to an increase in cesarean sections, prematurity, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

A major factor influencing IVF results is the gross inefficiency of human reproduction in vivo. A number of studies have attained the low fecundability of humans. Only 30% of normal women who attempt pregnancy in a given cycle are successful. There is a large loss of embryos prior to or around the time of implantation in vivo, and thus, many women have pregnancy losses without realizing that they have been pregnant because the menstrual period comes at the expected time. It can be concluded that many sperm and many oocytes, even from individuals who are fertile, do not have the ability to contribute to a normal pregnancy. This affects the success of both in vivo and in vitro fertilization.

The chance for success in successive IVF cycles decreases after 3 cycles. However, approximately 50% of women under age 35 have a live birth within 6 cycles of treatment. Most individuals find the emotional, physical, and financial consequences of going beyond 3 to 6 cycles too difficult, and adoption or donor embryos should be considered. Patients should also be aware that occasional spontaneous pregnancies occur even after failure of IVF.

Male infertility and IVF

The limited effectiveness of treatments for male infertility has provided a sizable number of individuals who desire IVF to overcome problems with sperm. Early experience with male factor infertility in IVF indicated that even placing sperm in the dish with the oocyte still left many sperm specimens with a handicap. Fertilization rates tended to be approximately one-half those achieved with normal sperm, and the pregnancy rates were correspondingly lower. On occasion, however, fertilization occurred with surprisingly few sperm available.

IVF provides the ability to visualize the results of sperm and egg interaction and thus to quickly determine if specific manipulations of the sperm can affect fertilization. A variety of sperm treatments have been attempted. One approach is to increase the number of sperm in the dish with the hope that even with abnormal specimens there will be a few normal sperm that can achieve fertilization. By increasing the numbers in each dish there will be more normal sperm per egg. A second approach is to isolate the best sperm from the specimen, not by the standard swim-up technique, but by using a variety of gradients. In some hands this has provided increased fertilization rates, but others have not found it to be a significant advantage. Similar contradictory results have been reported with drug treatment of the semen; the most popular such treatment has utilized pentoxifylline, which acts by increasing cyclic AMP in cells. The drug must be washed out from the sperm specimen before incubation with the egg because it may have adverse effects. Another treatment that has been used to enhance sperm is incubation in follicular fluid.

In men with sperm autoantibodies, in vitro fertilization is correlated with the extent the sperm are covered with antibodies.

Macromanipulation

The early techniques of making a hole in the zona to facilitate sperm entry and the injection of sperm through the zona into the perivitelline space have been superseded by the intracytoplasmic injection of single sperm (ICSI). The pioneering success of Van Steirteghem and his coworkers in Belgium has now been duplicated in many laboratories. The results have surpassed all expectations, with pregnancy rates equal to or slightly surpassing the normal fecundity rate.

Pregnancy rates are independent of any semen analysis parameter (number, motility, and morphology).

Sperm for injection have been obtained not only from the ejaculate, but also in azoospermia by microsurgery from the epididymis (MESA) or from the testis (TESE). Both percutaneous aspirations and open biopsies have been used. Fine-needle biopsy on occasion will locate small pockets of sperm or spermatids in otherwise hopeless appearing cases, such as the Sertoli cell only syndrome. Even the use of the sperm precursors in ICSI has produced successful pregnancies.

The success with spermatids and with immotile sperm, however, is substantially lower than with motile sperm. It is worthwhile, therefore, to vigorously search poor quality ejaculates for motile sperm.

ICSI is the treatment of choice for azoospermia and severe oligospermia (including the absence of sperm because of blockage in the ejaculatory ducts), and it is also indicated for genetic screening. Nevertheless, because of the relative infrequency of the various abnormalities, most IVF programs do not routinely kanotype prospective IVF patients. However, males with azoospermia or severe oligospermia should have genetic screening because their risk of genetic abnormalities is higher than men with normal sperm counts.

Concern arose that the gains achieved with ICSI, by bypassing the selection process exerted by the zona pellucida, might exact a price. The zona is a barrier to delivery, low birth weight, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

Perinatal morbidity and mortality increases as the number of fetuses increases. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

Prematurity, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

The multiple pregnancy rate is approximately 35% (30% twins, 5% triplets, and about 0.6% higher multiples). This leads to an increase in cesarean sections, prematurity, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

A major factor influencing IVF results is the gross inefficiency of human reproduction in vivo. A number of studies have attained the low fecundability of humans. Only 30% of normal women who attempt pregnancy in a given cycle are successful. There is a large loss of embryos prior to or around the time of implantation in vivo, and thus, many women have pregnancy losses without realizing that they have been pregnant because the menstrual period comes at the expected time. It can be concluded that many sperm and many oocytes, even from individuals who are fertile, do not have the ability to contribute to a normal pregnancy. This affects the success of both in vivo and in vitro fertilization.

The chance for success in successive IVF cycles decreases after 3 cycles. However, approximately 50% of women under age 35 have a live birth within 6 cycles of treatment. Most individuals find the emotional, physical, and financial consequences of going beyond 3 to 6 cycles too difficult, and adoption or donor embryos should be considered. Patients should also be aware that occasional spontaneous pregnancies occur even after failure of IVF.

Male infertility and IVF

The limited effectiveness of treatments for male infertility has provided a sizable number of individuals who desire IVF to overcome problems with sperm. Early experience with male factor infertility in IVF indicated that even placing sperm in the dish with the oocyte still left many sperm specimens with a handicap. Fertilization rates tended to be approximately one-half those achieved with normal sperm, and the pregnancy rates were correspondingly lower. On occasion, however, fertilization occurred with surprisingly few sperm available.

IVF provides the ability to visualize the results of sperm and egg interaction and thus to quickly determine if specific manipulations of the sperm can affect fertilization. A variety of sperm treatments have been attempted. One approach is to increase the number of sperm in the dish with the hope that even with abnormal specimens there will be a few normal sperm that can achieve fertilization. By increasing the numbers in each dish there will be more normal sperm per egg. A second approach is to isolate the best sperm from the specimen, not by the standard swim-up technique, but by using a variety of gradients. In some hands this has provided increased fertilization rates, but others have not found it to be a significant advantage. Similar contradictory results have been reported with drug treatment of the semen; the most popular such treatment has utilized pentoxifylline, which acts by increasing cyclic AMP in cells. The drug must be washed out from the sperm specimen before incubation with the egg because it may have adverse effects. Another treatment that has been used to enhance sperm is incubation in follicular fluid.

In men with sperm autoantibodies, in vitro fertilization is correlated with the extent the sperm are covered with antibodies.

Macromanipulation

The early techniques of making a hole in the zona to facilitate sperm entry and the injection of sperm through the zona into the perivitelline space have been superseded by the intracytoplasmic injection of single sperm (ICSI). The pioneering success of Van Steirteghem and his coworkers in Belgium has now been duplicated in many laboratories. The results have surpassed all expectations, with pregnancy rates equal to or slightly surpassing the normal fecundity rate.

Pregnancy rates are independent of any semen analysis parameter (number, motility, and morphology).

Sperm for injection have been obtained not only from the ejaculate, but also in azoospermia by microsurgery from the epididymis (MESA) or from the testis (TESE). Both percutaneous aspirations and open biopsies have been used. Fine-needle biopsy on occasion will locate small pockets of sperm or spermatids in otherwise hopeless appearing cases, such as the Sertoli cell only syndrome. Even the use of the sperm precursors in ICSI has produced successful pregnancies.

The success with spermatids and with immotile sperm, however, is substantially lower than with motile sperm. It is worthwhile, therefore, to vigorously search poor quality ejaculates for motile sperm.

ICSI is the treatment of choice for azoospermia and severe oligospermia (including the absence of sperm because of blockage in the ejaculatory ducts), and it is also indicated for genetic screening. Nevertheless, because of the relative infrequency of the various abnormalities, most IVF programs do not routinely kanotype prospective IVF patients. However, males with azoospermia or severe oligospermia should have genetic screening because their risk of genetic abnormalities is higher than men with normal sperm counts.

Concern arose that the gains achieved with ICSI, by bypassing the selection process exerted by the zona pellucida, might exact a price. The zona is a barrier to delivery, low birth weight, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

Prematurity, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.

The multiple pregnancy rate is approximately 35% (30% twins, 5% triplets, and about 0.6% higher multiples). This leads to an increase in cesarean sections, prematurity, and perinatal mortality. As noted previously, selective reduction can be offered as an option with multiples greater than two. This can be a difficult decision for some couples, but in view of the potential problems associated with a multiple pregnancy, it is an option deserving consideration.
with mild sperm impairment. The most frequent diagnosis is Klinefelter syndrome, which occurs in approximately 1 in every 500 males. Among other genetic abnormalities, azoospermia and hyponagonism can result because of a mutation that causes an abnormality in the beta-subunit of LH. An inactive mutation in the gene for the FSH receptor is associated with varying degrees of suppression of spermatogenesis, indicating that FSH is not absolutely essential for sperm production.

Approximately 1–2% of infertile males have congenital bilateral absence of the vas deferens. Most males with bilateral congenital absence of the vas deferens have at least two mutations in the cystic fibrosis transmembrane conductance regulator gene. Screening for cystic fibrosis mutations should be considered in males with bilateral absent vas deferens who are considering ICSI. A reasonable approach involves screening the female partner; the absence of cystic fibrosis mutations in the female reduces the risk of having a child with cystic fibrosis to a very low level.

Delayed mental development at one year of age has been reported in children born as the result of ICSI. 17% of ICSI children had mild or significantly delayed development (memory, problem solving, language skills) compared with 2% of IVF children and 1% of normal conception children. More reassuring are the reports of low rates of neurologic or developmental problems at 2 months of age, and normal mental development at age 2. With the large numbers of ICSI pregnancies, the long-term outcome of ICSI children will soon be apparent.

**IVF for Endometriosis**

Early experience with IVF in cases of endometriosis suggested that more severe disease exerted an adverse effect on oocyte recruitment, embryo development, and pregnancy rates. With greater experience, however, pregnancy rates (uninfluenced by the severity of the disease) have been achieved that are comparable to those with tubal disease.

**IVF for Autoimmunity**

ICSI is recommended for cases of sperm autoimmunity, and results are comparable to those obtained with other diagnoses. ICSI bypasses the zona, which is a major barrier to antibody-laden sperm.

**Other Techniques**

In addition to IVF, individuals with unexplained infertility have been offered a number of tactics to overcome hypothesized problems with gamete transport. For gamete intrafallopian transfer (GIFT), minilaparotomy or laparoscopy is used to aspirate oocytes following hyperstimulation of the ovary. Alternatively, aspiration can be performed more easily through the vagina, and laparoscopy is then used only for oocyte and sperm transfer (transfer the uterus produces poorer results). After the oocytes are identified in the laboratory, they are taken up into a transfer catheter that contains 100,000 sperm isolated by the swim-up technique. The transfer catheter is guided into the distal 4 cm of a fallopian tube and the contents gently discharged. Two oocytes can be placed in each tube although placement in one tube is more common and equally successful. Extra eggs obtained at retrieval can be fertilized and cryopreserved for future use. Success with GIFT in the collected statistics from the United States and Canada (1995) was 27.0% deliveries per retrieval. In the same year the success with IVF was 22.5% deliveries per retrieval. This difference is believed to be due largely to differences in patient selection. Ectopic pregnancy occurs in approximately 4.2% of GIFT pregnancies, somewhat higher than the 2.6% reported with IVF. The multiple pregnancy rate with GIFT is similar to that of IVF, approximately 35%.

In a variation of GIFT, called ZIFT, oocytes are obtained by vaginal aspiration, fertilized in vitro, and then 1 day later at the pronuclear stage placed in the fallopian tubes by the GIFT technique. The 1995 figures for ZIFT were 27.9% deliveries per retrieval. Whereas these figures continue to reflect an early impression that ZIFT was decidedly superior to IVF, this is not the case in current statistics. Indeed, prospective comparisons within the same clinics demonstrate similar results with uterine (IVF) versus tubal transfers (GIFT or ZIFT).

The choice of method, IVF, GIFT, or ZIFT, can be made on the basis of infertility factors, cost, and risk. If there is tubal damage, GIFT or ZIFT is not a good choice. In rare instances, where there is severe scarring of the cervix, one of the intratubal techniques is preferable. The comfort level of the program with each of the techniques and the program's experience in terms of pregnancy results are important considerations. Use of GIFT or ZIFT can incur greater costs, depending on the setting for the surgery and the anesthesia used.

Other attempts to overcome problems of gamete transport include injections of washed sperm and oocytes into the peritoneal cavity (POST) and cannulation of the fallopian tube via the cervix, as a conduit for injecting sperm directly into the tube. These methods have not gained wide usage. In a technique of limited applicability, because it risks pregnancy in a donor, a fertilized human ovm at the blastocyst stage can be removed by uterine lavage from a donor and placed transcrvically in the uterus of an infertile recipient.

**Ovum Donation**

A technique of proven value is ovum donation for older women, those with premature ovarian failure or women who are unresponsive to gonadotropin stimulation. The donor, who ideally should be under the age of 33 and has been both physically and psychologically screened, is down-regulated with a GnRH agonist, and then stimulated with gonadotropins as in IVF. Following retrieval, the donor's oocytes are fertilized with sperm from the recipient's partner. If the recipient is having periods, indication of endogenous hormone production, she is down-regulated with a GnRH agonist before starting on at least 10 days of estrogen (a common dose is estradiol, 4–8 mg per day). When ultrasonography indicates adequate endometrial development, progesterone is added to synchronize the endometrium with the egg retrieval. If pregnancy occurs, the hormone supplementation continues for at least 10 weeks, until placental steroidogenesis is firmly established. Success rates are lower in older recipients unless high doses of progesterone are provided for uterine support. A delivery rate in older women of approximately 40–50% per cycle can be achieved in a donor oocyte program. A cumulative birth rate can reach 96% with 4 or more cycles.

**Experimental Methods**

Oocytes can be matured in vitro and then cryopreserved. This technique would allow young women who require chemotherapy or pelvic irradiation to preserve a possibility for fertility in the future. However, fertilization rates and survival of cryopreservation is low for oocytes compared with the excellent survival obtained with cryopreserved embryos.

**Preimplantation Genetic Diagnosis**

Diagnosing genetic disorders before implantation provides couples with the option of foregoing the attempt to establish a pregnancy. This avoids the difficult decision whether or not to continue an affected pregnancy when the diagnosis is made at amniocentesis or by chorionic villus biopsy. There are 3 possible approaches for preimplantation diagnosis. The first is the removal of a polar body, but greater accuracy is obtained by studying both polar bodies. The polar body contains only one copy of the gene, but if the copy is found to be normal, it can be presumed that the oocyte contains a normal copy. However, this method is technically very difficult and subject to error if crossing-over occurs and both copies are present in the polar body. A second method is to biopsy cells that are destined to become the placenta. This requires culturing the embryo to the blastocyst stage, and then opening the zona pellucida in the 5–6 day embryo. The disadvantage is the lower pregnancy rate when the embryo is transferred at this later stage. The third method is the removal of a single cell from the 6–8 cell embryo (blastomere biopsy) for DNA amplification by polymerase chain reaction and analysis with fluorescent probes specific for chromosomes 13,18, 21, and the X and Y chromosomes. The biopsy procedure does not affect development and implantation, and the diagnostic testing is rapid; the biopsy and DNA analysis are accomplished within 8 hours.

Preimplantation genetic diagnosis has successfully detected single gene defects in disorders such as cystic fibrosis, Duchenne’s muscular dystrophy, sickle cell disease, hemophilia. Tay-Sachs disease, Lesch-Nyan syndrome, and at least 11 other diseases. These methods can also be used to determine the sex of the embryo for couples who are at risk for transmitting X-linked disorders.

The utilization of molecular biology techniques for preimplantation genetic diagnosis is associated with some significant risks. Polymerase chain reaction amplification does not always succeed. An erroneous diagnosis can occur because of contamination by DNA from surrounding cells or from sperm that are attached to the zona


Hypogonadism caused by a single amino acid substitution in the b-subunit of luteinizing hormone, Fertil Steril 59:1, 1993.


Human pregnancy following oocyte and sperm transfer to the uterus, Lancet i:1031, 1998.


Methods of early diagnosis represent the best and most rewarding approach. Patients who present because of acute symptoms (frequently in emergency rooms) are assessed. Standard descriptions in texts are based upon older reports and, thus, older methods of diagnosis. Suspicion of the diagnosis and rapid recourse to the methods would be effective, for the first time in 1883, deliberately and successfully performed a laparotomy to ligate the broad ligament and a ruptured tube.

By 1885, Tait had accumulated a relatively large number of successful cases. Asepsis, anesthesia, and antibiotics (and blood transfusions) combined to save the lives of many women. But diagnosis was still difficult, and surgical intervention was relatively late. Even in the first half of the 20th century, the maternal mortality rate in the United States ranged from 200 to 400 per 10,000 cases of ectopic pregnancy. As dramatic as the contribution of immediate salpingectomy coupled with simultaneous blood transfusion was, progress in the last 20 years has been even more impressive. Treatment has shifted from the saving of lives to the preservation of fertility.

The Centers for Disease Control first began to report the incidence of ectopic pregnancies in the U.S. in 1970. In 1970, there were 17,800 ectopic pregnancies, and by 1992, the number had increased to 108,800, a rate increase of 4.5 to 19.7 ectopic pregnancies per 1000 pregnancies. However, at the same time, the fatality rate decreased from 35.5 to 3.4 per 10,000 ectopic pregnancies, a decrease of 90%.

The increase in ectopic pregnancies has not been paralleled by a similar increase in sexually transmitted diseases (STDs), and, therefore, the increased incidence of ectopic pregnancies is not due to STDs alone. Ectopic pregnancies do occur in totally normal tubes, suggesting that abnormalities of the conceptus or maternal hormonal changes can function as etiologic factors. The other important contributing factors are reconstructive tubal surgery, assisted reproductive technologies, and, most importantly, earlier and more accurate diagnosis.

Today, management intervention occurs prior to tubal rupture in more than 80% of cases. This can be attributed directly to three diagnostic advances: a highly specific and sensitive immunoassay for human chorionic gonadotropin (HCG), ultrasonography, and the use of laparoscopy.

Etiology and Clinical Presentation

Even though the risk of death from ectopic pregnancy has declined dramatically, ectopic pregnancy is the second leading cause of overall maternal mortality in the United States and the leading cause of pregnancy-related deaths during the first trimester. This represents a combination of a lack of access to appropriate services and misdiagnosis.

Ectopic pregnancy is the great masquerader. The clinical presentation can vary from vaginal spotting to vasomotor shock with hematoperitoneum. The classic triad of delayed menses, irregular vaginal bleeding, and abdominal pain is most commonly not encountered. The exact frequency of clinical symptoms and signs is hard to assess. Standard descriptions in texts are based upon older reports and, thus, older methods of diagnosis. Suspicion of the diagnosis and rapid recourse to the methods of early diagnosis represent the best and most rewarding approach. Patients who present because of acute symptoms (frequently in emergency rooms) are
usually at a more advanced gestational age compared to asymptomatic infertility patients being followed closely because of their increased risk for ectopic pregnancy.

Differential Diagnosis

- Normal intrauterine pregnancy.
- Ruptured ovarian cyst.
- Bleeding corpus luteum.
- Spontaneous miscarriage.
- Salpingitis.
- Appendicitis.
- Adnexal torsion.
- Endometriosis.
- Diverticulitis.

Relevant factors in a patient’s medical history include prior tubal surgery, the use of assisted reproductive technology, exposure to diethylstilbestrol (DES), previous pelvic inflammatory disease, vaginal douching, and the method of contraception. In addition, cigarette smoking is associated with a 2-fold increased risk for ectopic pregnancy.2 However, most patients presenting with an ectopic pregnancy do not have a recognized risk factor, suggesting dysfunctional problems in tubal transport or impaired implantation due to some abnormality in the conceptus. Nevertheless, pregnancies following tubal surgery or treatment with one of the methods of assisted reproduction should, ideally, be diagnosed immediately and followed closely with HCG titers and ultrasonography.

Previous Pelvic Inflammatory Disease

In Westrom’s classic report, women with a history of salpingitis (verified by laparoscopy) had a four-fold increased risk of ectopic pregnancy.3 Salpingitis damages the endosalpinx, resulting in agglutination of the mucosal folds and adhesion formation. The risk of an ectopic pregnancy increases with each inflammatory episode. Evidence of chlamydial infection (circulating antibodies) is associated with a greater than two-fold increased risk of ectopic pregnancy.4 A similar increased risk is associated with douching, but the presence of infection may be the reason for the douching.

Prior Tubal Surgery

Women who have had tubal surgery have an increased risk of ectopic pregnancy. High-risk surgery includes any infertility surgery on the tube, but not abdominal or pelvic surgery that avoids the tubes.5 Women with an ectopic pregnancy treated by conservative surgery have a ten-fold increased risk of a subsequent ectopic. Ectopic pregnancies occur after tubal occlusion procedures for sterilization; the risk with postpartum sterilization is very low, comparable to that observed in oral contraceptive users.6 8 9 10 11 12 With interval sterilization, bipolar tubal coagulation is more likely to result in ectopic pregnancy than is mechanical occlusion.13 14 15 This is attributed to fistula formation that allows sperm passage, and this may explain the difference in ectopic pregnancy rates between interval and postpartum sterilization because postpartum procedures are mostly by the Pomeroy method. Ectopic pregnancies following tubal ligation usually occur two or more years after the sterilization, rather than immediately after. In the first year after sterilization, about 6% of sterilization failures will be ectopic pregnancies, but the majority of pregnancies that occur 2–3 years after occlusion will be ectopic.16 Overall, the ectopic risk in women with interval sterilizations is 80% less than that in nonsterilized women; however, the relative risk is 3.7 times that of women using oral contraception and 2.8 times that with barrier methods of contraception.17 About one-third of pregnancies that occur after tubal sterilization are ectopic.18

The Use of Assisted Reproductive Technology

Pregnancies occurring simultaneously in different body sites (heterotopic pregnancies) are a rare condition, occurring in 1 of 30,000 spontaneous pregnancies. The incidence of combined pregnancy among patients who have undergone one of the assisted reproduction procedures (in vitro fertilization, gamete intrafallopian transfer, and even superovulation) is much higher, closer to 1 in 100 pregnancies.19 20 21 22 23 Close monitoring of pregnancies in these programs is important to prevent a deleterious delay in the treatment of an ectopic pregnancy.

Is the increased risk with these treatment methods due to ovulation induction (with superovulation and elevated levels of hormone influence on the tubes) or due to previous tubal disease? A case-control study has concluded that the risk of ectopic pregnancy was increased four-fold with ovulation induction, but not further increased when ovulation induction was used for in vitro fertilization.24 This would indicate that the multiple eggs and high hormone levels are the important factors. Retrograde embryo migration is believed to be a major mechanism.

Method of Contraception

The risk of ectopic pregnancy is reduced with all methods of contraception except the progesterone-containing intrauterine device.25 26 27

The IUD has been traditionally listed as a risk factor for ectopic pregnancy. The previous use of an IUD does not increase the risk of a subsequent ectopic pregnancy.28 And it should be emphasized that the current use of modern copper-bearing IUDs does NOT increase the risk of ectopic pregnancy and, in fact, offers considerable protection.29 30 31 32 33 34 35 The largest study, a World Health Organization multicenter study, concluded that IUD users were 50% less likely to have an ectopic pregnancy when compared to women using no contraception.36 However, if an IUD user becomes pregnant, the pregnancy is more likely to be ectopic. About 3–4% of IUD pregnancies have been ectopic, making the actual occurrence a rare event.

Ectopic Pregnancy Rates per 1000 Woman-Years37 38 39 40

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tubal surgery</td>
<td>20.0</td>
</tr>
<tr>
<td>Previous ectopic</td>
<td>10.0</td>
</tr>
<tr>
<td>Previous salpingitis</td>
<td>4.0</td>
</tr>
<tr>
<td>Assisted reproduction</td>
<td>4.0</td>
</tr>
<tr>
<td>Age &lt; 25</td>
<td>3.0</td>
</tr>
<tr>
<td>Previous pelvic infection</td>
<td>3.0</td>
</tr>
<tr>
<td>Infertility</td>
<td>2.5</td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td>2.5</td>
</tr>
<tr>
<td>Vaginal douching</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The lowest ectopic pregnancy rates are seen with the most effective IUDs, like the TCu-380A (90% less likely compared to noncontraceptors).41 The rate is about...
one-tenth the ectopic pregnancy rate associated with the Lippes Loop or with devices with less copper such as the TCu-200. The progesterone-releasing IUD has a higher rate, probably because its action is limited to a local effect on the endometrium. While very few ectopic pregnancies have been reported with the levonorgestrel IUD, 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.

The risk of ectopic pregnancy does not increase with increasing duration of use with the TCU-380A or the levonorgestrel IUD. In a 7-year prospective study, not a single ectopic pregnancy was encountered with the levonorgestrel IUD. In 8000 woman-years of experience in randomized multicenter trials, there has been only a single ectopic pregnancy reported with the TCU-380A (which is one-tenth the rate with the Lippes Loop or TCu-200).

The protection against ectopic pregnancy provided by the TCU-380A and the levonorgestrel IUD makes these IUDs acceptable choices for contraception in women with previous ectopic pregnancies.

The risk of an ectopic pregnancy during use of Norplant is lower than the general rate. However, because of the impressive contraceptive efficacy of Norplant, when pregnancy does occur, ectopic pregnancy should be suspected. With the progestin-only minipill, ectopic pregnancy is not prevented as effectively as intrauterine pregnancy. Although the overall incidence is not increased, the situation is similar to that with Norplant. When pregnancy occurs, an ectopic gestation must be suspected.

Ectopic Sites

Almost all ectopic pregnancies are located in the tube. Although relatively uncommon, ectopic pregnancies in nontubal sites are very susceptible to complications, especially hemorrhage. For example, abdominal pregnancies are often misdiagnosed, and the mortality rate is 17 times greater compared to the overall ectopic rate. Contrary to the experience with tubal pregnancy, recurrence in nontubal sites is rare.

### Sites of Ectopic Implantation

<table>
<thead>
<tr>
<th>Site</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fallopian tube:</td>
<td>80 %</td>
</tr>
<tr>
<td>Ampullary segment</td>
<td></td>
</tr>
<tr>
<td>Isthmic segment</td>
<td>12 %</td>
</tr>
<tr>
<td>Fimbrial end</td>
<td>5 %</td>
</tr>
<tr>
<td>Cornual and interstitial</td>
<td>2 %</td>
</tr>
<tr>
<td>Abdominal</td>
<td>1.4%</td>
</tr>
<tr>
<td>Ovarian</td>
<td>0.2%</td>
</tr>
<tr>
<td>Cervical</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

The Methods of Early Diagnosis

Patients with normal intrauterine pregnancies can present with the same symptoms encountered in patients with unruptured ectopic pregnancies. The best way to diagnose ectopic pregnancy is to be highly suspicious and sensitive to its possibility, and to utilize the new tools of diagnosis: the quantitative measurement of b-HCG and ultrasonography. Laparoscopy is necessary only when the diagnosis is in doubt, or when laparoscopy is the technique selected for surgical treatment.

The Quantitative Measurement of HCG

HCG is secreted by the syncytiotrophoblast and reaches a maximal level of 50,000–100,000 IU/L at 8–10 weeks of gestation. The maternal circulating HCG concentration is approximately 100 IU/L at the time of the expected but missed menses. Virtually 100% (but not absolutely all) of patients suspected of an ectopic pregnancy, but not pregnant, will have a negative blood assay for b-HCG. Contrast this present day sensitivity of the HCG assay with the urinary tests of the past. In the 1960s and early 1970s, the urinary pregnancy tests were positive in only 50% of patients with ectopic pregnancies.

The modern assay for HCG measures the serum level of the HCG beta subunit. b-HCG is the largest gonadotropin b-subunit, containing a larger carbohydrate moiety and 145 amino acid residues, including a unique carboxyl terminal tail piece of 24 amino acid groups. It is this unique part of the HCG structure which allows the production of highly specific antibodies and the utilization of highly specific immunologic assays. In some laboratories, the lower limit of the assay is 4–6 IU/L, but even that level is almost never associated with a false negative. With a detection limit of less than 5 IU/L for the serum b-HCG assay, there should be no false-negative results; however, this situation (very rarely) can be encountered. When the clinical picture is confusing, a definitive diagnosis by laparoscopy is warranted.

A landmark observation was reported from Yale in 1981, documenting that HCG levels approximately double every 2 days in early, normal intrauterine pregnancies, and that a lesser increase is associated with ectopic pregnancies and spontaneous miscarriages. In the first 6 weeks of normal pregnancy, the concentration of HCG in the maternal blood follows a well-recognized pattern. The rate of increase is nonlinear, changing with advancing gestational age and increasing HCG concentrations. However, during the time period when the diagnosis of ectopic pregnancy is most important, from 2 to 4 weeks after ovulation, the relationship between HCG levels and gestational age is linear, approximately doubling every 2 days until the titer is greater than 10,000. Use of the HCG titer requires medical judgment. Some ectopic pregnancies will display a normal rise in titer (at least for awhile), and some normal pregnancies (about 10%) will have an abnormal doubling time.

The Clinical Usefulness of the Quantitative Measurement of b-HCG

1. **Assessment of pregnancy viability.** Most of the time, but not always, a normal rate of rise indicates a normal pregnancy. Clinical decisions require serial measurements of HCG.
2. **Correlation with ultrasonography.** When the titer exceeds 1000–1500 IU/L, vaginal ultrasonography should identify the presence of an intrauterine gestation. With multiple gestation, a gestational sac will not be apparent until the titer is a little higher. In an asymptomatic patient, repeat ultrasonography 2–3 days later is warranted.
3. **Assessment of treatment results.** Declining levels are consistent with effective medical or surgical treatment. Persistent or rising levels indicate the presence of viable trophoblastic tissue.

Vaginal Ultrasonography

The difficulty in diagnosis is establishing the cause of abnormal HCG levels: is it an ectopic pregnancy or a spontaneous miscarriage? The addition of ultrasonography has made an important contribution to this differential diagnosis. A “gestational sac” is an ultrasonographic landmark, located beneath the endometrial surface, and visualized by transvaginal ultrasonography beginning 30–35 days after the last menstrual period. It is a sonolucent center surrounded by a thick echogenic rim, formed by a decidual reaction round the chorionic sac. The yoke sac is the first structure visualized within the gestational sac, at 5 weeks after the last menstrual period. Cardiac motion is first observed at 5.5–6 weeks gestation.

The discriminatory zone is that HCG titer above which a gestational sac can be identified with ultrasonography. Previously, with abdominal ultrasonography, this level was approximately 6000 IU/L. Transvaginal ultrasonography has now established the discriminatory zone at a level that varies from 1000 to 1500 IU/L. This level is achieved approximately one week after the time of expected menses. Institutions must establish their own exact discriminatory zone based on the sensitivity of their immun assay and their equipment and the ability of their ultrasonographers. Keep in mind that in a multiple pregnancy, the discriminatory zone will be a little higher, requiring an extra 2–3 days for a gestational sac to become visible. Thus, an accurate gestational age is very helpful; failure to demonstrate the presence of a gestational sac 24 or more days after conception (38 or more gestational days) usually indicates an ectopic pregnancy.
Along with the increase in incidence of ectopic pregnancy associated with superovulation, dizygotic twinning is more common, and thus a combined intrauterine pregnancy and extraterine pregnancy (heterotopic pregnancy) is more frequent. Heterotopic pregnancy should especially be considered when the pregnancy is the result of one of the methods of assisted reproductive technology.

Color and pulsed Doppler increases the sensitivity of vaginal ultrasonography. This method adds physiologic information to the anatomic picture of regular ultrasonography. A small intrauterine gestational sac, without the double sac sign (decidua and membranes) or the presence of a yolk sac is hard to distinguish from the pseudosac of an ectopic pregnancy. A pseudogestational sac due to endometrial bleeding occurs in about 10% of ectopic pregnancies. Local vascular changes associated with a true gestational sac differentiate an intrauterine pregnancy from the pseudosac of ectopic pregnancy. Vascular pulses increase with pregnancy and produce a “warm” appearance with color Doppler ultrasonography. In addition, high-velocity arterial flow is detected even with very early pregnancies.

A failed intrauterine pregnancy can be associated with either elevated flow velocity around the trophoblast or very low velocities. Thus, a characteristic peritrophoblastic arterial flow correlates with gestational sac size and HCG levels. Ectopic masses can also be distinguished by the surrounding abnormal color mapping. Doppler ultrasound, therefore, has greater sensitivity and is technically less challenging (and quicker). Although traditional ultrasonography can reveal an adnexal mass, Doppler flow imaging can indicate that the mass is an ectopic pregnancy by documenting abnormal vascular activity of the mass combined with the relatively cool uterine vasculature. This difference between Doppler and standard ultrasonography is greatest in early pregnancy, and thus this greater accuracy can allow earlier initiation of medical treatment.

### Important Observations With Vaginal Ultrasonography

1. **Documentation of an intrauterine sac.**
   - An experienced ultrasonographer should be able to identify a viable intrauterine pregnancy if the HCG titer is 1000–1500 IU/L or greater. An HCG titer of greater than 1000–1500 IU/L with no intrauterine sac is consistent with an ectopic pregnancy.

2. **Adnexal masses.**
   - An ectopic equal to or greater than 2 cm in diameter should be identified by ultrasonography. Vaginal ultrasonography should also establish large ectopics that require different consideration. Currently an ectopic pregnancy with a diameter of 4 cm or more is a relative contraindication to medical treatment.

3. **Adnexal cardiac activity.**
   - The presence of cardiac activity in an ectopic pregnancy, detectable when the HCG titer is approximately 15,000–20,000 IU/L, is a relative contraindication to medical treatment.

### The Progesterone Level

Single serum progesterone levels have a wide spectrum with considerable overlap between normal and ectopic pregnancies. This measurement must be viewed as an adjunct to HCG levels and ultrasonography. The concentration of the serum progesterone is usually lower in ectopic pregnancies. A value of 25 ng/mL or more is 98% of the time associated with a normal intrauterine pregnancy, while a value of less than 5 ng/mL identifies a nonviable pregnancy, regardless of location.

The value of the serum progesterone is to help make a decision regarding the viability of a possible intrauterine pregnancy prior to curettage. In most cases, however, this is a decision easily made by the combined results of the clinical presentation, the HCG titers, and ultrasonography. The great majority of patients will have a progesterone level between 10 and 20 ng/mL at presentation, significantly limiting the clinical usefulness of progesterone measurement.

The value of 25 ng/mL as an indicator of a normal intrauterine pregnancy was established in women with spontaneous ovulations and pregnancies. The appropriate number for women receiving medication for the induction of ovulation is probably higher, and in these cases, the use of the progesterone level is even more limited.

### Uterine Curettage

The purpose of uterine curettage is to determine the presence or absence of villi to rule out a nonviable intrauterine pregnancy. In most cases, curettage is helpful when a serum progesterone is less than 5 ng/mL and a non-rising HCG titer is below 1000 IU/L. Curettage with examination of the curettings can be used to prevent unnecessary laparoscopies in patients undergoing spontaneous miscarriages. Floating the curettings in saline will usually identify villi if present, but not always. The saline flotation of curettings has been reported to be incorrect in 6.6% of patients with ectopic pregnancies and incorrect in 11.3% of patients with intrauterine pregnancies. Because of this inaccuracy, permanent sections and follow-up HCG titers are necessary for confirmation.

### Culdocentesis

Culdocentesis to seek the presence of unclotted blood was for a period of time a valuable technique to aid in the diagnosis of an ectopic pregnancy. We have progressed to the point where the relative accuracy of culdocentesis is no longer sufficient. Furthermore, the presence of blood in the cul-de-sac does not mean that an ectopic pregnancy has ruptured. Therefore, a positive culdocentesis is of no help in deciding whether to treat medically or surgically. Culdocentesis no longer has a place in the differential diagnosis of ectopic pregnancy.

### The Treatment of Ectopic Pregnancy

#### Expectant Management

Part of the increased incidence of ectopic pregnancy is due to earlier diagnosis detecting ectopic pregnancies, which previously resolved and remained clinically undiagnosed. Not all tubal pregnancies progress to clinical manifestations, and, therefore, expectant management of ectopic pregnancies diagnosed very early is an appropriate choice. Expectant management includes the monitoring of clinical symptoms, HCG titers, and ultrasonography findings. Approximately one-fourth of women presenting with ectopic pregnancy can be managed expectantly, and 70% of this select group of patients will avoid surgery and experience successful outcomes. Success with expectant management decreases with increasing HCG levels; best results are achieved when the initial HCG level is less than 2000 IU/L. The long-term outcome (subsequent intrauterine and ectopic pregnancies) is similar to that with active treatment interventions. The following criteria are reasonable requirements for expectant management:
Criteria for Expectant Management

1. Falling HCG titer.
2. Ectopic pregnancy definitely in the tube.
3. No significant bleeding.
4. No evidence of rupture.
5. Ectopic mass not larger than 4 cm in greatest diameter.

Medical Treatment

Medical treatment of unruptured ectopic pregnancies is appealing for several reasons: less tubal damage, less cost, and, hopefully, enhanced potential for future fertility. Methotrexate, a folic acid antagonist that interferes with DNA synthesis, has a long history of effectiveness against trophoblastic tissue, derived from experience in the treatment of hydatidiform moles and choriocarcinoma. Methotrexate was used in the 1960s to treat the difficult problem of trophoblastic tissue left in situ after removal of an abdominal pregnancy. Methotrexate was first used to treat an ectopic pregnancy in Japan in 1982. The first U. S. experience was reported by Ory in 1986. The guidelines for the safe and effective use of methotrexate were established by Stovall and his colleagues.

Criteria for Patient Selection

1. The patient is healthy, hemodynamically stable, reliable, and compliant.
2. Ultrasonography should fail to find an intrauterine pregnancy, and uterine curettage should fail to obtain villi.
3. The ectopic pregnancy measures 4 cm or less in its greatest diameter.
4. There is no evidence of rupture of the ectopic pregnancy.
5. HCG titers greater than 10,000 IU/L and fetal cardiac activity are relative contraindications. However, even patients with fetal cardiac activity have been successfully treated.

Prior to Methotrexate Treatment

1. Administer Rhogam if patient is Rh-negative and greater than 8 weeks gestation.
2. Obtain baseline liver and renal function tests, complete blood and platelet counts.
3. Consider uterine curettage.

Patient Instructions

The following are avoided until HCG titers are negative: alcohol use, sexual intercourse, and the use of folic acid-containing vitamins.

The Multiple Dose Method. The initial protocols for treatment utilized multiple doses of methotrexate together with citrovorum factor (folinic acid) to minimize side effects. Treatment with this method has been reported to be 70–94% successful. Failures have been more common with HCG levels greater than 5000 IU/mL, and thus the presence of fetal cardiac activity is generally a contraindication. Side effects (in 3–4% of patients) include mild stomatitis, gastritis, diarrhea, and transient elevations in liver enzymes. Significant reactions (bone marrow suppression, dermatitis, pneumonitis) have been very rare. The incidence of nonresponders and/or tubal rupture is 3–4%.

In our own experience, the onset of abdominal cramping occurring 3–4 days after the initiation of methotrexate treatment (in approximately 60% of patients) produces some anxious moments. The concern, of course, is that the ectopic pregnancy is rupturing. Although this is always a possibility, the cramping is usually a side effect of the methotrexate and resolves in a day or two. Occasionally, hospitalization is necessary until hemodynamic stability is verified.

Multiple Dose Methotrexate Protocol

Treatment is discontinued when a decline is observed in two consecutive daily HCG titers, or after 4 doses of methotrexate.

| Day 1: Baseline studies. Methotrexate | 1.0 mg/kg im |
| Day 2: Citrovorum factor | 0.1 mg/kg im |
| Day 3: Methotrexate | 1.0 mg/kg im |
| Day 4: Citrovorum factor HCG titer | 0.1 mg/kg im |
| Day 5: Methotrexate HCG titer | 1.0 mg/kg im |
| Day 6: Citrovorum factor HCG titer | 0.1 mg/kg im |
| Day 7: Methotrexate HCG titer | 1.0 mg/kg im |
| Day 8: Citrovorum factor HCG titer Complete blood and platelet counts. Renal and liver function tests. | 0.1 mg/kg im |

Weekly: HCG titer until negative.

Approximately 20% of patients will require only one dose of methotrexate, and 20% will require 4 doses. The side effects are encountered in those patients who require multiple doses. Changes in blood counts and liver enzymes are so infrequent and so mild, that daily monitoring is unnecessary.

In terms of subsequent fertility, methotrexate compares favorably with conservative laparoscopic surgery, and the rate of subsequent ectopic pregnancy is lower. In a randomized trial comparing multiple dose methotrexate with laparoscopy, the results were as follows:

<table>
<thead>
<tr>
<th>Treatment for persistent trophoblast</th>
<th>Multiple Dose Methotrexate</th>
<th>Laparoscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>No further treatment</td>
<td>82%</td>
<td>72%</td>
</tr>
<tr>
<td>Treatment for persistent trophoblast</td>
<td>4%</td>
<td>20%</td>
</tr>
</tbody>
</table>

The ultrasonographic picture of a mass persists after HCG titers become negative. The time for resolution of the mass is variable, and often it takes several months. Thus, the persistence of a mass should not be interpreted as a treatment failure.
The Single Dose Method. Experience with the multiple dose method indicated that a significant number of patients responded promptly and did not require several doses. With lesser dosing, fewer side effects could be anticipated, and the use of citrovorum factor could be abandoned. The results with a single dose are very good (80–90%), even with very high HCG titers and the presence of fetal cardiac activity. The HCG titers usually keep rising for 3 days after treatment but by day 7 are declining. Full resolution requires 3 to as much as 6 weeks. Serious side effects are virtually absent. If there is less than a 15% decline on day 7 (the usual assay variation), the treatment protocol is repeated (necessary in approximately 8% of patients).

**Single Dose Methotrexate Protocol**

<table>
<thead>
<tr>
<th>Day</th>
<th>Baseline studies.</th>
<th>Methotrexate</th>
<th>HCG titer.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td></td>
<td>50 mg M/F im.</td>
<td></td>
</tr>
<tr>
<td>Day 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 7</td>
<td>HCG titer.</td>
<td></td>
<td>Complete blood and platelet count.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Liver and renal function tests.</td>
</tr>
<tr>
<td>Weekly</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Oral Methotrexate. Experience with the oral method of treatment is not encouraging. One comparison with placebo in the expectant management of ectopic pregnancy found no indication of any effect.

**Important Cautions**

The medical treatment of ectopic pregnancy requires compulsive compliance. An ectopic pregnancy can exist in the absence of detectable HCG. Although in this instance, the extratubal pregnancy is usually degenerating and associated with an indolent clinical course, rupture can still occur. Clinicians should always be alert for the possibility of rupture (3–4% of medically treated cases). A satisfying decline in HCG titers does not guarantee against rupture. The average time for HCG to return to nondetectable levels is about 4 weeks.

The risk of tubal rupture is about 10% when the HCG titer is less than 1000 IU/L, and if the ectopic is isthmic, a risk of rupture is still present with a titer of 100 or less. Remember that a negative HCG assay means that HCG is not present in levels greater than the sensitivity of the assay; therefore, trophoblastic tissue can still be present, secreting minimal amounts of HCG (below the limits of the assay). Ectopic pregnancies will continue to adhere to their historical record; always expect behavior that is an exception to the general rule.

Although experience is limited with the medical treatment of relatively large masses (with a fetus present and high levels of HCG), case reports indicate a greater risk of bleeding and problems. Carefully selected cases with HCG titers greater than 10,000 IU/L might warrant medical treatment, but as a general rule these cases deserve surgical treatment.

**Special Indications for Methotrexate.** Treatment with methotrexate is especially useful when the pregnancy is located in a site (cervix, ovary, or cornua) where surgical treatment carries significant risk. Methotrexate treatment is an attractive option when an ectopic pregnancy is in the interstitial portion of the tube or growing in the wall of the uterus (diagnosed by ultrasonography).

**Salpingocentesis**

Salpingocentesis is the injection of a substance directly into the gestational sac within the tube, either at laparoscopy or under ultrasound guidance. Various substances have been used, including methotrexate, potassium chloride, prostaglandins, and hyperosmotic glucose. The efficacy, safety, and the long-term impact on fertility have not been established. Thus far local injections have been associated with inconsistent results; at least one clinical trial was discontinued because of poor results with tubal injection of methotrexate while another claimed excellent results, especially when the HCG level was under 5000 IU/L. Circulating levels of methotrexate are similar when gestational sac injection is compared to intramuscular injection. Thus, local treatment with methotrexate offers no obvious advantage over systemic treatment. Hyperosmotic glucose (a 50% solution) appears to be safe and effective when the HCG titers are less than 2500 IU/L.

**Surgical Treatment**

With earlier diagnosis, conservative surgery to preserve fertility has replaced the life-saving procedure of salpingectomy. Linear salpingostomy along the antimesenteric border to remove the products of conception is the procedure of choice for ectopic pregnancies in the ampullary portion of the tube. Ectopic pregnancies in the ampulla are usually located between the lumen and the serosa, and thus these are ideal candidates for linear salpingostomy. Segmental excision with either simultaneous or delayed microsurgical anastomosis is the preferred procedure for isthmic pregnancies. Although linear salpingostomy is possible for a small and unruptured gestation, isthmic pregnancies reflect a damaged endosalpinx, and these patients do poorly with linear salpingostomy (with a high rate of recurrent ectopic pregnancy).

Occasionally, an ampullary pregnancy can be expressed through the fimbrial end of the tube (milking the tube), but this procedure is associated with a higher incidence of persistent and recurrent ectopic pregnancy, undoubtedly due to invasion of the tube by the trophoblastic tissue. However, fimbrial expression of an ectopic pregnancy that is easily dislodged is acceptable. Interstitial pregnancy at the utero-tubal junction usually requires surgical excision, and even hysterectomy if bleeding cannot be controlled. The first unruptured ectopic pregnancy treated with methotrexate was an interstitial pregnancy, and this is now the treatment of choice if diagnosis is achieved early enough.

Patients with compromised fertility do better when the tube that contains the ectopic pregnancy is conserved (even when the opposite tube appears to be normal). However, in patients with a history positive for previous tubal disease, the risk of a recurrent ectopic pregnancy in the same tube is very much higher, and in this case, some argue in favor of salpingectomy. When performing a salpingectomy, a cornual wedge excision as prophylaxis against recanalulation and a subsequent ectopic pregnancy is no longer considered to be necessary. An effort should be made to retain both ovaries when appropriate as a resource for the future use of in vitro fertilization.

**Indications for Salpingectomy**

- Childbearing completed.
- Second ectopic pregnancy in the same tube.
- Uncontrolled bleeding.
- Severely damaged tube.

These procedures can be accomplished either by laparotomy or laparoscopy. The choice of surgical method and specific procedure is determined by the patient’s condition, desire for future fertility, the location, size, and state of the ectopic pregnancy, and the experience of the surgeon. The relative contraindications to laparoscopy include extensive pelvic adhesions, hematoperitoneum, and an ectopic pregnancy greater than 4 cm diameter. Hemodynamic instability is an absolute contraindication.

Linear salpingostomy through the laparoscope achieves results comparable with those obtained at laparotomy. Hemostasis is the key and several methods are used, including the use of vasopressin, microcautery, and laser. The gains are notable: outpatient versus inpatient cost and a more rapid recovery. Almost all patients can now be successfully treated with conservative surgery. Comparisons of the different types of surgery (laparoscopic salpingostomy, laparotomy salpingostomy, and laparotomy salpingectomy) indicate that the surgical technique chosen is less important in determining future fertility than the causes of the ectopic pregnancy.
Treatment of an Ectopic Pregnancy After Tubal Ligation

An ectopic pregnancy after a previous tubal ligation is usually located in the segment of tube containing the fimbria. The pregnancy occurs because of small channel recanalization through the ligature site, allowing sperm to migrate toward the ovary. A prophylactic procedure should be highly considered. Removing both fimbrial segments and fulgurating the proximal segments (either by laparoscopy or laparotomy) will prevent the recurrence of another ectopic pregnancy.

Treatment of Persistent Trophoblastic Tissue

The risk of a persistent ectopic pregnancy with conservative surgery by laparotomy is 5%. Laparoscopic salpingostomy is associated with a higher rate of persistent trophoblastic tissue; approximately 15% of patients will require further treatment. Persistence of ectopic trophoblastic tissue can be associated with hemorrhage and tubal rupture (usually within 2 weeks); however, regression without clinical sequelae is the general rule. For this reason, weekly HCG measurements are necessary following surgery. The incidence of persistent trophoblastic tissue is greater (not surprising) with higher HCG liters, and relatively rare with a liter less than 3000 IU/L. The risk of persistent trophoblastic tissue is very significant with a hematosalpinx greater than 6 cm in diameter, an HCG liter greater than 20,000 IU/L, and a hematoperitoneum greater than 2000 mL. Rupture is unlikely for an ampullary pregnancy with an HCG level of 100 or less, but not so for an isthmic pregnancy.

The average time for HCG levels to become undetectable is 4 weeks, but it can take 6 weeks. The need for treatment of persistent trophoblastic tissue can emerge in a few days or not until 1 month later. Although reoperation is always a treatment option, the use of methotrexate is preferable. Prophylactic treatment can be administered with a single dose of methotrexate (1mg/kg), or a single dose of methotrexate (15 mg/m²) can be administered after diagnosis.

Low and declining HCG levels warrant only close surveillance; only persistent or rising liters require treatment (a small minority of patients). Symptomatic patients, of course, usually demand surgical therapy.

The problem of persistent trophoblastic tissue after surgery makes diagnosis of an ectopic pregnancy even more important. With sufficiently early diagnosis, medical treatment becomes the method of choice.

Rh Sensitization

Despite underutilization of Rhogam in Rh-negative women, no apparent increase in sensitization has been observed. This indicates that ectopic pregnancies do not contain sufficiently large quantities of fetal red blood cells. The use of Rhogam should be considered only for ectopic pregnancies that are older than 8 weeks gestation.

Fertility After Ectopic Pregnancy

Having had one ectopic pregnancy, a woman is at increased risk for another, especially if infectious pathology is present. Nevertheless, after one ectopic pregnancy, the chance of having a live-borning infant is 85% with the next pregnancy. After 2 ectopic pregnancies, the risk of ectopic pregnancy is increased nearly 10-fold, and consideration should be given to in vitro fertilization.

Chapter References

### Laboratory Values for Selected Measurements in Urine

<table>
<thead>
<tr>
<th>Substance</th>
<th>Conventional Units</th>
<th>Conversion Factor</th>
<th>SI Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortisol, free</td>
<td>10–90 µg/24 hr</td>
<td>2.759</td>
<td>28–250 nmol/24 hr</td>
</tr>
<tr>
<td>Estrogens, total</td>
<td>5–25 µg/24 hr</td>
<td>3.67</td>
<td>18–92 nmol/24 hr</td>
</tr>
<tr>
<td>17-Hydroxycorticosteroids</td>
<td>2–6 mg/24 hr</td>
<td>2.759</td>
<td>5.5–15.5 µmol/24 hr</td>
</tr>
<tr>
<td>17-Ketosteroids</td>
<td>6.0–15 mEq/24 hr</td>
<td>3.467</td>
<td>21–52.5 µmol/24 hr</td>
</tr>
</tbody>
</table>

### SI Prefixes and Their Symbols

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>giga</td>
<td>G</td>
</tr>
<tr>
<td>mega</td>
<td>M</td>
</tr>
<tr>
<td>kilo</td>
<td>k</td>
</tr>
<tr>
<td>hecto</td>
<td>h</td>
</tr>
<tr>
<td>deka</td>
<td>da</td>
</tr>
<tr>
<td>deci</td>
<td>d</td>
</tr>
<tr>
<td>centi</td>
<td>c</td>
</tr>
<tr>
<td>milli</td>
<td>m</td>
</tr>
<tr>
<td>micro</td>
<td>µ</td>
</tr>
<tr>
<td>nano</td>
<td>n</td>
</tr>
<tr>
<td>pico</td>
<td>p</td>
</tr>
<tr>
<td>femto</td>
<td>f</td>
</tr>
<tr>
<td>atto</td>
<td>a</td>
</tr>
</tbody>
</table>

### Laboratory Values for Selected Measurements in Blood, Plasma, and Serum

<table>
<thead>
<tr>
<th>Substance</th>
<th>Conventional Units</th>
<th>Conversion Factor</th>
<th>SI Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTH, adrenocorticotropin hormone</td>
<td>10–80 pg/mL</td>
<td>0.2202</td>
<td>2.2–17.6 pmol/L</td>
</tr>
<tr>
<td></td>
<td>&lt;50 pg/mL</td>
<td>0.2202</td>
<td>&lt;11 pmol/L</td>
</tr>
<tr>
<td>Androstenedione</td>
<td>60–300 ng/dL</td>
<td>0.0349</td>
<td>2.1–10.5 nmol/L</td>
</tr>
<tr>
<td>Calcium, total</td>
<td>8.5–10.5 mg/dL</td>
<td>0.25</td>
<td>2.1–2.6 mmol/L</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>&lt;200 mg/dL</td>
<td>0.0259</td>
<td>&lt;5.2 mmol/L</td>
</tr>
<tr>
<td>LDL-cholesterol</td>
<td>60–130 mg/dL</td>
<td>0.0259</td>
<td>1.6–3.4 mmol/L</td>
</tr>
<tr>
<td>HDL-cholesterol</td>
<td>30–70 mg/dL</td>
<td>0.0259</td>
<td>0.8–1.8 mmol/L</td>
</tr>
<tr>
<td>Cortisol</td>
<td>5–25 µg/dL</td>
<td>27.6</td>
<td>140–690 nmol/L</td>
</tr>
<tr>
<td></td>
<td>3–12 µg/dL</td>
<td>27.6</td>
<td>80–330 nmol/L</td>
</tr>
<tr>
<td></td>
<td>&lt;50% of AM value</td>
<td>27.6</td>
<td>&lt;50% of AM value</td>
</tr>
<tr>
<td>DHAS</td>
<td>80–350 µg/dL</td>
<td>0.0027</td>
<td>2.2–9.5 pmol/L</td>
</tr>
<tr>
<td>Estradiol</td>
<td>0.05–0.25 µg/dL</td>
<td>28.86</td>
<td>1.5–7.3 pmol/L</td>
</tr>
<tr>
<td>Estrone</td>
<td>2–10 ng/mL</td>
<td>30.3</td>
<td>60–300 pmol/L</td>
</tr>
<tr>
<td>Estrone</td>
<td>20–400 pg/mL</td>
<td>3.67</td>
<td>70–1500 pmol/L</td>
</tr>
<tr>
<td>FSH, reproductive years</td>
<td>&lt;10 ng/mL</td>
<td>1.0</td>
<td>&lt;10 µg/L</td>
</tr>
<tr>
<td>Glucose, fasting</td>
<td>70–110 mg/dL</td>
<td>0.0556</td>
<td>4.0–6.0 mmol/L</td>
</tr>
<tr>
<td>Growth hormone</td>
<td>&lt;10 ng/mL</td>
<td>1.0</td>
<td>&lt;10 µg/L</td>
</tr>
<tr>
<td>Thyroid stimulating hormone</td>
<td>0.35–5.0 µU/mL</td>
<td>1.0</td>
<td>5–20 IU/L</td>
</tr>
<tr>
<td>Cortisol</td>
<td>100–300 ng/dL</td>
<td>0.03</td>
<td>3–9 nmol/L</td>
</tr>
<tr>
<td>Insulin, fasting</td>
<td>5–20 µU/mL</td>
<td>7.175</td>
<td>35–145 pmol/L</td>
</tr>
<tr>
<td>Insulin-like growth factor-I</td>
<td>0.3–2.2 U/mL</td>
<td>1000</td>
<td>300–2200 IU/L</td>
</tr>
<tr>
<td>LH, reproductive years</td>
<td>5–20 mIU/mL</td>
<td>1.0</td>
<td>5–20 IU/L</td>
</tr>
<tr>
<td>Progesterone</td>
<td>&lt;3 ng/mL</td>
<td>3.18</td>
<td>&lt;9.5 nmol/L</td>
</tr>
<tr>
<td>Follicular phase</td>
<td>5–30 ng/mL</td>
<td>3.18</td>
<td>16–95 nmol/L</td>
</tr>
<tr>
<td>Secretory phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prolactin</td>
<td>1–20 ng/mL</td>
<td>44.4</td>
<td>44.4–888 pmol/L</td>
</tr>
<tr>
<td>Testosterone, total</td>
<td>20–80 ng/mL</td>
<td>0.0347</td>
<td>0.7–2.8 nmol/L</td>
</tr>
<tr>
<td>Testosterone, free</td>
<td>100–200 pg/mL</td>
<td>0.0347</td>
<td>35–700 pmol/L</td>
</tr>
<tr>
<td>TSH, thyroid stimulating hormone</td>
<td>0.35–5.0 µU/mL</td>
<td>1.0</td>
<td>0.35–5.0 mU/L</td>
</tr>
<tr>
<td>Test</td>
<td>Normal Range</td>
<td>Value</td>
<td>Reference Range</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------</td>
<td>-------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Thyroxine, free T₄</td>
<td>0.8–2.3 ng/dL</td>
<td>1.29</td>
<td>10–30 nmol/L</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>40–250 mg/dL</td>
<td>0.0113</td>
<td>0.5–2.8 mmol/L</td>
</tr>
<tr>
<td>Triiodothyronine, T₃, total</td>
<td>80–220 ng/dL</td>
<td>0.0154</td>
<td>1.2–3.4 nmol/L</td>
</tr>
<tr>
<td>Triiodothyronine, T₃, free</td>
<td>0.13–0.55 ng/dL</td>
<td>15.4</td>
<td>2.0–6.5 pmol/L</td>
</tr>
<tr>
<td>Triiodothyronine, reverse</td>
<td>8–35 ng/dL</td>
<td>15.4</td>
<td>120–540 pmol/L</td>
</tr>
</tbody>
</table>