CLINICAL OBSTETRICS AND GYNECOLOGY
*Volume 55, Number 3, 692–705*


# *The Modern Infertility Evaluation*

**MARC A. FRITZ, MD**

*Department of Obstetrics and Gynecology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*

**Abstract:** The modern diagnostic evaluation of the infertile couple reflects a growing reliance on assisted reproductive technologies and the trend toward a more evidence-based medical practice. The recommended evaluation no longer includes some of the traditional diagnostic tests, applies other tests more selectively, and includes a new test that helps to define a couple's prognosis and best choice of treatment. All tests are easily performed, allowing clinicians to complete a basic but still thorough evaluation quickly and easily.
**Key words:** infertility, semen analysis, ovulation, hysterosalpingography, ovarian reserve

The diagnostic evaluation of the infertile couple has changed significantly in recent years. Current methods reflect advances in, and our growing reliance on assisted reproductive technologies, and the trend toward a more evidence-based medical practice. The purpose of this review is to describe the components of the modern infertility evaluation, focusing on how the traditional evaluation has changed and on the evidence that signaled the need for change.

The traditional evaluation of infertility was designed to separate the reproductive process into its component parts and to test the integrity of each in efforts to identify the limiting "factor(s)." The evaluation of ovulatory function typically involved basal body temperature (BBT) recordings or timed serum progesterone determinations. Cervical factors were assessed with the postcoital test (PCT), which involved the collection of midcycle cervical mucus within hours after intercourse and microscopic examination to evaluate the quality and quantity of mucus and characteristics of the sperm-mucus interaction. Evaluation of the male focused on semen analysis to identify men with abnormally low sperm counts or motility. Uterine factors were evaluated by hysterosalpingography (HSG) to identify congenital or acquired morphologic abnormalities of the uterine cavity, including mullerian anomalies, endometrial polyps, myomas, and synechiae; endometrial biopsy and histologic dating were performed to exclude luteal phase deficiency (LPD) predisposing to implantation failure. HSG also was the primary method for evaluating tubal patency, with laparoscopy and chromotubation being the obvious alternative and also the preferred method for exclusion of peritoneal factors, referring primarily to endometriosis and adnexal adhesions.

The modern infertility evaluation has the same basic design and purpose and

*Correspondence: Marc A. Fritz, MD, Department of Obstetrics and Gynecology, UNC School of Medicine, CB#7570, 4001 Old Clinic Building, Chapel Hill, NC 27599-7570.*

*The author declares that there is nothing to disclose.*

uses many of the same methods. However, it no longer includes some of the traditional diagnostic tests, because they have been proven invalid or yield results having little or no impact on treatment decisions. It applies other tests more selectively, and includes some new elements that help to define a couple's prognosis and best choice of treatment.

The evaluation of infertility should focus on the couple and not on one or the other partner. Consequently, both partners should be encouraged to attend each visit during evaluation, whenever possible. Joint visits help to ensure that both partners understand the rationale for recommended tests and procedures, and have the chance to have their questions answered directly. Counseling should be an ongoing process with regular visits to discuss test results and their implications and to ensure that all of the couple's medical, emotional, and financial concerns are addressed in a timely and effective manner.

## *Initial Consultation*

Evaluation generally should be offered to all couples who have failed to conceive after a year or more of regular intercourse, by which time approximately 85% of couples attempting pregnancy will succeed. Earlier evaluation is warranted for couples wherein the male partner has known or suspected poor semen quality or the female partner has irregular or infrequent menses, a history of pelvic infection or endometriosis, or is over 35 years of age.[1] The initial consultation with an infertile couple should include careful medical histories, provide education on normal reproductive efficiency, describe recommended screening tests, and explain the major known causes of infertility, and the methods for their detection.

Relevant medical history includes any previous pregnancies and associated complications, cycle length and characteristics, coital frequency and sexual dysfunction, the duration of infertility and results of any previous evaluation or treatment, previous pelvic infections or surgery, current medications, use of tobacco, alcohol and recreational drugs, and symptoms of thyroid disease, pelvic pain, dyspareunia, galactorrhea, or hirsutism. Relevant physical findings include weight and body mass index, palpable thyroid abnormalities, breast secretions, signs of androgen excess, vaginal or cervical abnormalities, and pelvic tenderness, organ enlargement or mass.

Many, indeed most, couples are surprised to learn that human reproduction is not nearly as efficient as they believed. It is useful to inform them that normally fertile couples having regular intercourse will conceive at a rate of approximately 20% per month, that virtually all pregnancies result from intercourse occurring sometime within the 6-day interval ending on the day of ovulation, and that even when intercourse occurs on the very day of ovulation, the likelihood of achieving pregnancy is no greater than approximately 35%.[2] That perspective is especially helpful when discussing the prognosis for success with different treatment options, which usually is expressed in terms of cycle fecundity (the probability of achieving a successful pregnancy per cycle) and ranges between 5% and 50%.

Screening for cystic fibrosis is recommended for couples wherein one or both partners is white or of Ashkenazi Jewish descent and should be offered to all patients upon request; sequential screening (testing one partner, and the second only if the first is identified as a carrier) is the most cost-effective approach. Other genetic screening should be guided by family history and ethnicity, in accordance with existing guidelines. Women without previous documented rubella or varicella infection or vaccination should be screened for immunity. Screening for sexually transmitted infections is recommended for women at moderate-to-high risk for

infection and should consider that the Centers for Disease Control recommends screening for chlamydia and gonorrhea (nucleic acid-based tests), syphilis (rapid plasma reagin), hepatitis B (hepatitis B surface antigen), and voluntary screening for human immunodeficiency virus type I (HIV-1) for all pregnant women at the first prenatal visit. For women receiving donor sperm, the American Society for Reproductive Medicine recommends mandatory screening for HIV-1 and 2, hepatitis B, hepatitis C (hepatitis C antibody), cytomegalovirus, and human T-cell lymphocyte virus types I and II, and suggests additional screening for chlamydia and gonorrhea. For male partners of women receiving donor sperm, the American Society for Reproductive Medicine strongly recommends screening for HIV-1, hepatitis B and C, and syphilis.

## *The Infertility Evaluation*

A basic infertility evaluation should include tests aimed at the 4 most important causes of infertility: (1) ovulatory dysfunction, (2) abnormalities of semen, (3) abnormalities of the uterus and fallopian tubes, and (4) reproductive aging. Among infertile couples, the primary cause relates to ovulatory dysfunction in approximately 20%, to male factors in 30%, and to tubal and pelvic pathology in another 30%, with most of the rest remaining unexplained. When the medical histories and physical examination do not point clearly to a specific cause for infertility, the best approach is to begin with tests of ovulation and semen quality, deferring invasive diagnostic procedures in the female partner until those are completed.

### OVULATORY DYSFUNCTION

Available tests for ovulation include BBT monitoring, serum progesterone measurements, urinary luteinizing hormone (LH) monitoring (ovulation predictor kits), and serial transvaginal ultrasonography. However, often the menstrual history alone is sufficient for diagnosis of ovulatory dysfunction. Menses in normally ovulating women typically are regular, consistent in duration and flow characteristics, and are preceded by a recognizable pattern of molimina. Conversely, anovulatory women exhibit irregular or infrequent menses, which vary in duration and character, and have no consistent molimina. When the menstrual history clearly indicates ovulatory dysfunction, no specific diagnostic tests are necessary.

#### *BBT*

Although traditional BBT monitoring (body temperature taken each morning upon awakening, before arising, plotted on graph paper) is still useful, its use has declined steadily in popularity, primarily because the method is subjective, often difficult to interpret confidently, and can become tedious over time as it begins each day with a reminder of reproductive failure. The primary advantages that BBT monitoring has over other methods for evaluating ovulation are its low cost and the ability to reveal a grossly short luteal phase (a subtle form of ovulatory dysfunction) that otherwise might not be recognized.

#### *Serum Progesterone*

Serum progesterone measurements are a simple and objective measure of ovulatory function, as long as they are appropriately timed. Levels generally remain $< 1$ ng/mL during the follicular phase, increase slightly at the time of ovulation, rise steadily thereafter to peak approximately one week after ovulation, and then decline progressively over the week preceding the onset of menses. Any concentration exceeding 3 ng/mL provides reliable, qualitative evidence of recent ovulation.[3] The test does not reveal when ovulation occurred; it indicates only that it did. Although a cycle day 21 serum progesterone level is

common in practice, it is not necessarily the best time to obtain the test in all women. The normal ovulatory cycle is 25 to 35 days in duration and has a 13- to 15-day luteal phase. Therefore, women may ovulate normally as early as cycle day 10, or as late as day 22. Day 21 is a perfect choice for women with 28-day cycles, but a poor choice for those with 35-day cycles. The best time to obtain a serum progesterone concentration will vary with the length of the menstrual cycle, aiming for approximately one week before the expected onset of menses.

Although serum progesterone levels also have been used to evaluate the quality of luteal function, a truly accurate measure requires daily testing across the luteal phase, which is neither cost-effective nor practical. Less frequent sampling, regardless of how well timed, cannot reliably define the quality of luteal function for two reasons. First, progesterone concentrations in normal and abnormal cycles and in conception and nonconception cycles overlap greatly. Second, progesterone is secreted by the corpus luteum in a distinctly pulsatile manner and concentrations can vary up to 8-fold within an interval of only hours.[4]

***Ovulation Predictor Kits***

Commercial "ovulation predictor kits" are designed to detect the preovulatory LH surge in urine and typically become "positive" when concentrations exceed a threshold value normally observed only during the LH surge. Reliable use requires that testing be performed daily, usually beginning 2 to 3 days before the surge is expected, based on the overall length of the cycle. Although the test can be positive for more than one day, the first positive test best predicts the time of ovulation and further testing is unnecessary. Although early morning would seem the most logical time to test because urine typically is most concentrated then, results correlate best with the LH surge in serum when the test is performed during the late afternoon or evening (04:00 to 10:00 PM). Testing can yield false-negative or false-positive results in 5% to 10% of cycles. Accuracy varies among the different kits available, with the best products predicting ovulation within the following 24 to 48 hours with greater than 90% probability. As ovulation usually occurs 14 to 26 hours after detection of the surge and the oocyte is fertilizable only for approximately 24 hours, the interval of greatest fertility spans the day of the surge and the next two days, with the day after surge detection being the one best day for timed intercourse or insemination. Ovulation predictor kits invite women to become actively involved in their care and can accurately define the length of the luteal phase, which is a crude but still useful measure of luteal function. However, their greatest advantage over other tests of ovulation is their ability to predict when ovulation will occur, in advance, which is particularly helpful for couples having intercourse infrequently or requiring insemination.

***Ultrasonography***

Serial transvaginal ultrasonography can reveal the size and number of preovulatory follicles, provide the most accurate estimate of the time of ovulation, and are essential to the safety and effectiveness of ovulation induction with exogenous gonadotropins. Serial ultrasonography can document progressive growth of the preovulatory follicle, followed by its collapse, the loss of distinct margins and an increase in internal echoes, and an increase in the volume of fluid in the cul-de-sac. Serial ultrasonography can be very useful for timing indicated inseminations when LH monitoring proves unreliable or unsuccessful, but their cost and logistical demands otherwise are difficult to justify. Serial ultrasonography has been advocated as the best means to detect subtler forms of ovulatory dysfunction, such as the "luteinized unruptured follicle

syndrome,'' but such abnormalities generally are rare and there is little evidence to indicate that they occur more often or consistently in infertile women than in fertile women.

## MALE FACTORS

Semen analysis has been, and still is, the primary method for initial evaluation of the male partner. In men with abnormal semen parameters, further endocrine, urologic, and genetic evaluation may be indicated.

### *Semen Analysis*

Traditional reference standards for semen quality are based on comparisons of values observed in male partners of fertile and infertile couples, without specific exclusion of female infertility factors and, therefore, do not necessarily represent the average ranges observed in fertile men. More importantly, they also do not represent the minimum values consistent with fertility.

Two studies have helped to better define threshold values that can differentiate fertile from subfertile men. The first, conducted by the Reproductive Medicine Network (RMN; a group of fertility centers selected by the National Institutes of Health) found that a sperm density (concentration) $>48$ million/mL, progressive motility (the proportion of sperm exhibiting purposeful forward motion) $>63\%$, and morphology (the proportion of sperm having normal shape using ''strict criteria,'' as described below) $>12\%$ normal predicted fertility. Conversely, a density $<13.5$ million/mL, progressive motility $<32\%$, and morphology $<9\%$ normal predicted subfertility; intermediate values were indeterminate.[5] The RMN study observed that the odds of male infertility increased with the number of parameters in the subfertile range; the probability increased 2- to 3-fold when 1 was abnormal, 5- to 7-fold when 2 were abnormal, and 16-fold when all 3 were abnormal. The second and more recent study, conducted by the World Health Organization (WHO), defined lower reference limits for semen analyses, representing the fifth centile in a population of almost 2000 men from 8 countries on 3 continents whose partners conceived within 12 months. The lower reference limits for fertile men were 15 million sperm/mL, 32% progressive motility, and 4% normal morphology (''strict criteria'').[6] The normal reference limits suggested by the WHO compare closely to those suggested by the earlier RMN study. Together, these data provide reliable, clinically relevant values for use in the evaluation of infertile men.

Sperm morphology as judged by the ''strict'' criteria originally described by Kruger and adopted by the WHO in 1999, is the best current predictor of sperm function, implying their ability to attach to, penetrate, and fertilize an oocyte. Very low values ($<4\%$ normal) predict poor or failed fertilization in cycles of in vitro fertilization (IVF) and are widely recognized as an indication for applying intracytoplasmic sperm injection (ICSI), rather than conventional fertilization methods in such cycles. Not surprisingly, strict morphology also correlates with cycle fecundity after intrauterine insemination (IUI); the probability of pregnancy rises with the proportion of normal sperm and generally is poor when values are $<4\%$ normal.[7] Consequently, strict morphology probably is most relevant for couples with a mild oligospermia (decreased sperm concentration), or asthenospermia (decreased sperm motility), or unexplained infertility, because they are the logical candidates for IUI, with or without ovarian stimulation. Unfortunately, strict morphology generally is not available outside of specialty andrology laboratories that usually are closely affiliated with IVF programs because the technique requires specialized training and relatively high test volume to maintain proficiency. The older WHO standards for sperm morphology still used

widely in hospital laboratories have no demonstrable predictive value.

***Endocrine Evaluation***

As recommended by the American Urological Association, men with oligospermia (particularly when the sperm concentration is < 10 million/mL) and those with sexual dysfunction (decreased libido, impotence) merit further endocrine evaluation, beginning with measurement of the serum follicle-stimulating hormone (FSH) and total testosterone.[8] Those with abnormally low testosterone concentrations, < 300 ng/dL, require additional evaluation by measuring the serum free testosterone, LH, and prolactin levels. In men with abnormal spermatogenesis, the FSH level is normal or high and LH and testosterone levels are normal. In those with testicular failure, the FSH and LH concentrations are high and the testosterone level is low or normal. In men with hypogonadotropic hypogonadism, FSH, LH, and testosterone levels are all low and, as in women, magnetic resonance imaging is indicated to detect hypothalamic and pituitary lesions.

***Urologic Evaluation***

Formal urologic evaluation is recommended for men with subfertile semen parameters. Physical examination focuses on testis volume, the extent of virilization, the epididymus and vas deferens, the presence of any palpable varicocele, and a digital rectal examination that might reveal evidence of active inflammation or obstruction. In men with severe oligospermia (sperm concentration < 5 million/mL) or azoospermia (no sperm seen), palpable vasa, low-volume ejaculates, and normal testis volume, transrectal ultrasonography is indicated to detect ejaculatory duct obstruction, particularly when the ejaculate has an acidic pH and contains little or no detectable fructose (the secretions of the seminal vesicles are alkaline and contain fructose).[8] Suspected obstructions can be confirmed by aspiration of the seminal vesicles or vesiculography. Any aspirated sperm can be cryopreserved for later ICSI and transurethral ejaculatory duct resection can restore normal semen quality in some men. Testis biopsy generally is reserved for azoospermic men having a palpable vas deferens, normal testis volume, and normal gonadotropin concentrations. If the testicular tissue yields sperm, they can be cryopreserved for ICSI. Histologic evidence of normal spermatogenesis implies obstruction at some level, which may then warrant surgical exploration, with or without vasography, to define the location.

***Genetic Evaluation***

Additional genetic evaluation should be offered to men with severe oligospermia or azoospermia who may be considering IUI or ICSI.[8] Cystic fibrosis screening is indicated for all men with unilateral or bilateral vasal or epididymal agenesis or obstruction. Mutations of the cystic fibrosis transmembrane conductance regulator (CFTR) gene are highly associated with congenital bilateral absence of the vas deferens (CBAVD) and related abnormalities. Almost all men with cystic fibrosis have CBAVD and at least two thirds of men with CBAVD have a detectable CFTR mutation. As the association is so strong, all men with CBAVD generally should be assumed to have a mutation, even when screening fails to detect one. Screening and genetic counseling also is indicated for their female partners to ensure the couple is well informed about the risks for transmitting cystic fibrosis or CBAVD to their offspring.

For men with nonobstructive azoospermia or severe oligospermia, karyotyping should be offered to detect chromosomal aneuploidy (eg, 47,XXY Klinefelter syndrome), translocations, or inversions; their prevalence is 7% overall and inversely related to sperm density.[8] Additional specific genetic testing to detect microdeletions in the "azoospermia factor" (AZF) region of the Y chromosome (not detected by

karyotype) also should be offered, because they have a similar prevalence and will be transmitted to any male offspring (thereby predisposing to abnormal spermatogenesis and subfertility).[8]

## CERVICAL FACTOR

The PCT was long regarded as one of the essential components of the basic infertility evaluation. Its clinical value was based on two arguments. The first was that the test identified couples with abnormalities in the quality or quantity of cervical mucus production or sperm-mucus interaction who could benefit from IUI. The second was that results predicted the probability of treatment-independent pregnancy, with those having an abnormal test being 2 to 3 times less likely to conceive without treatment.[9] However, over time, those arguments have been dismantled and lost relevance.

Studies examining the efficacy of IUI for treatment of "cervical factor infertility" have suggested only a very modest benefit, at best, compared with no treatment. Any prognostic value the PCT might have is largely limited to young couples with unexplained infertility of relatively brief duration, because the test cannot be expected to predict outcomes for anovulatory women or those with tubal occlusive disease, and significant male factors are better identified by semen analysis. Expectant management is reasonable for some couples with unexplained infertility, because pregnancies generally occur no more frequently with treatment, only sooner, but for those desiring more than one child and for older women, sooner is arguably also better. Moreover, even young couples desiring a single child most often will reject expectant management.

Only one randomized controlled trial has examined the efficacy of the PCT.[10] A substantial number of "new" infertile couples were randomized to undergo evaluation, with or without the PCT. In those having a PCT, an abnormal test (absent or nonmotile sperm) was judged valid only if the cervical mucus was of good quality or if a second test confirmed the result, taking care to ensure it was performed immediately before ovulation, based on serial transvaginal ultrasonography. Evaluation and treatment of both groups otherwise proceeded in a standard manner. After 24 months, the cumulative pregnancy rate among the women evaluated with the PCT was significantly lower than in those who had no PCT, and pregnancy rates in women having normal and abnormal tests were not different.[10]

In sum, routine PCT is no longer recommended, for several reasons. First, abnormalities of mucus or sperm-mucus interaction are rarely the primary cause of infertility, and the two most common causes of a "cervical factor"—chronic cervicitis and cervical stenosis—can be identified by careful speculum examination. Second, the test is inconvenient and embarrassing for many women. Third, the PCT has no standardized methodology or interpretation and has poor reproducibility, even in specialty centers. Fourth, neither the results of the PCT nor treatment for an abnormal test have any measurable impact on outcomes. Lastly, results of the PCT seldom affect treatment, because contemporary treatments for unexplained infertility—IUI with or without ovarian stimulation and IVF—both "treat" any unrecognized cervical factor by bypassing the cervix altogether.

## UTERINE FACTORS

Uterine factors include abnormalities of the uterine cavity and the disorder known as LPD, long regarded as a cause of both infertility and recurrent pregnancy loss.

### *Abnormalities of the Uterine Cavity*

HSG has been, and remains, the most common test for evaluation of the uterine

cavity. The test readily detects common developmental anomalies of the uterus, such as a septate or bicornuate uterus, although further imaging (3-dimensional transvaginal ultrasonography or magnetic resonance) is required to reliably distinguish the two. HSG also detects other important acquired uterine abnormalities, including intrauterine adhesions, submucous myomas, and most, but not all, significant endometrial polyps. As a bonus, HSG provides useful information regarding tubal patency.

Sonohysterography, also known as saline sonography, is a reasonable alternative to HSG for evaluation of the uterine cavity, particularly for women having no risk factors for tubal disease and for those in whom tubal status is largely irrelevant, as in couples with severe male factor infertility in whom IVF/ICSI is the only reasonable option. Moreover, sonohysterography is preferred over HSG for women having symptoms suggesting a cavity mass lesion. For example, in women with menorrhagia (regular, heavy, or prolonged menses), the test has approximately 50% positive predictive value (PPV) and 90% negative predictive value (NPV), meaning that it can suggest lesions that are not real, but reliably excludes a mass lesion.[11] Like HSG, sonohysterography also has a bonus in that it can reveal otherwise unsuspected ovarian pathology, such as an endometrioma.

Diagnostic hysteroscopy is the gold standard among methods for evaluating the uterine cavity, but generally offers few advantages over sonohysterography. Chief among these is that hysteroscopy has greater specificity than sonohysterography, because it distinguishes between endometrial polyps and submucous myomas, which sonohysterography often cannot, but that differentiation also is largely unimportant in the evaluation of infertile women because the finding of any mass lesion generally leads to operative hysteroscopy for excision.

### *LPD*

LPD is a disorder classically characterized by a deficiency of corpus luteum progesterone production, in amount, duration, or both. Most have viewed it as a subtle form of ovulatory dysfunction, with inadequate follicular development resulting in decreased luteal cell mass and steroidogenic capacity, although any hypothalamic/pituitary dysfunction during the luteal phase also could deprive the corpus luteum of essential LH support.

The basic pathophysiological concept of LPD is that inadequate progesterone secretion causes delayed secretory endometrial maturation. At least in theory, a severe delay predisposes to infertility caused by failed implantation, due to an unreceptive endometrium and/or an aging embryo, and a mild delay predisposes to early pregnancy loss, caused by a tardy hCG "rescue" signal (arriving after the corpus luteum already has started to regress) that cannot stimulate or maintain the level of progesterone secretion required to support early pregnancy.

For more than 50 years, the gold standard among methods for diagnosis of LPD was endometrial "dating," a technique described originally in a classic 1950 article, which appeared in the inaugural issue of the journal *Fertility and Sterility* and remains among the most often cited in all of gynecology.[12] Endometrial dating was based on the histologic features of the glands and stroma in tissue specimens obtained during the late luteal phase. In brief summary, an "expected date" defined by the day of sampling, counting back from the onset of the next menses (designated cycle day 28 or luteal day 14), was compared with an assigned "histologic date," with those more than two days "out of phase" considered abnormally "delayed." The diagnosis of LPD required the finding be observed in two separate, preferably consecutive, cycles.

Many challenged the validity of histologic endometrial dating over the years.

Some researchers questioned how a "normal" standard could be based on analysis of tissues obtained from infertile women (as it was). Others argued that interpretations based on the onset of the menses ignored normal variations in luteal phase duration and the effect that biopsy had on when menses began or was perceived to start. Still others observed that the technique had wide intraobserver and interobserver variations that affected interpretation.

In 2004, endometrial histologic dating was critically examined in two separate studies. The first recruited a large group of normally cycling women of reproductive age with proven fertility, each randomly assigned to have endometrial biopsy on luteal day 1 to 14, as defined by detection of the urinary LH surge (designated luteal day 0). Consecutive sections of each tissue specimen were examined by three independent expert gynecologic pathologists, who were masked to the day of sampling, and each tissue specimen was scored on 32 distinct histologic features, using criteria the group of pathologists had agreed upon before the study began.[13] The study found that although the secretory endometrium does indeed exhibit a consistent sequence of maturational changes, the classical features used for histologic dating are much less temporally discrete than was originally described, vary significantly between cycles and individuals and, consequently, cannot reliably define a specific luteal day, or even a narrow interval of days. The second study examined the prevalence of "delayed" endometrial maturation in matched groups of fertile and infertile women, observing the "abnormality" more frequently in fertile women than in infertile women.[14]

Not surprisingly, endometrial histologic dating is no longer recommended for the evaluation infertility or recurrent early pregnancy loss. Endometrial histology does not reliably reflect the amount or duration of progesterone secretion, varies widely between individuals, between cycles within individuals, and among observers, and lacks the accuracy and precision required for a valid diagnostic test, primarily because it cannot discriminate fertile from infertile women.

## TUBAL FACTOR

The methods available for evaluating the fallopian tubes include traditional HSG, laparoscopic "chromotubation," and the chlamydia antibody test (CAT).

### *HSG*

HSG has been, and remains the most useful basic test of tubal patency. Over the years, controversy has raged over the relative advantages and disadvantages of water-soluble and oil-soluble contrast media. Advocates of water-soluble contrast have argued that oil obscures internal tubal architectural detail and disperses poorly, reducing the test's ability to detect adnexal adhesions (as suggested by observations of contrast loculation), and carries unique risks of granuloma formation and embolism. Those who favor oil-based contrast argue those risks are rare and are persuaded by evidence suggesting that fertility is improved after HSG when an oil-based contrast medium is used. However, a 2009 Cochrane systematic review including two trials comparing oil-soluble and water-soluble media found no evidence of a significant difference in the odds of pregnancy.[15] Consequently, both water-soluble and oil-soluble contrasts are appropriate, depending on preference.

As a test of tubal patency, HSG is approximately 60% sensitive and 95% specific, meaning that when it suggests obstruction, the tubes are often truly patent, but when it demonstrates patency, the tubes are almost always truly open.[16] The relatively poor sensitivity of HSG as a test of tubal patency results from the difference in test accuracy for diagnosis of proximal and distal tubal occlusion. The diagnosis of distal tubal obstruction

generally is accurate, but apparent proximal tubal occlusions are often not real, representing artifacts of transient uterine contractions, so-called "tubal spasm," or catheter placement (with the tip lying near one tubal orifice). The HSG diagnosis of proximal tubal obstruction must, therefore, be confirmed, either by repeating the study, or by performing either fluoroscopic or hysteroscopic selective tubal catheterization.

### *Laparoscopic "Chromotubation"*

The obvious alternative to HSG for the evaluation of tubal patency is laparoscopic "chromotubation," involving the introduction of a dilute solution of methylene blue or indigo carmine through the cervix during laparoscopy, and observing its movement through the fallopian tube. Although generally more accurate than HSG, the diagnosis of proximal tubal occlusion has the same pitfalls as HSG and, ideally, should be confirmed by selective tubal catheterization.

### *CATs*

CATs also have some clinical utility for the detection of tubal pathology. The CAT is based on the detection of immunoglobulin-G antibodies resulting from infection with *Chlamydia trachomatis*. Several European studies have suggested that the sensitivity of CAT for detection of tubal pathology approaches that of HSG and laparoscopic chromotubation.[17] At least in theory, the CAT should help to identify women with tubal pathology who might benefit most from more specific tests, such as HSG or laparoscopic chromotubation. However, at present, the diagnostic accuracy of the CAT has not been established and the test is not used widely in the United States.

## PERITONEAL FACTORS

Peritoneal factors relating to infertility include endometriosis and adnexal adhesions resulting from previous pelvic surgery or infection. Either may distort pelvic anatomy and interfere with ovum capture or transport. Diagnostic laparoscopy was long considered a routine and essential element of the infertility evaluation, primarily for detection of peritoneal factors that otherwise might escape detection, but the best available evidence indicates that routine diagnostic laparoscopy is neither justified nor cost effective.

### *Endometriosis*

The results of two randomized trials with similar designs, one conducted in Canada and the other in Italy, indicate that the diagnosis and treatment of minimal or mild endometriosis has relatively little importance. In both studies, laparoscopy was performed in groups of women with unexplained infertility, and those having minimal and mild endometriosis were randomized to treatment by excision or ablation of disease, or to no treatment; the patients were then followed for up to a year without further intervention. Although more pregnancies were observed in treated women than in untreated women, the differences were small. Analysis of combined data from both studies found that the "number needed to treat" (the calculated inverse of the treatment effect difference) was 12, meaning that one had to find and treat 12 women with minimal and mild endometriosis to achieve 1 more pregnancy than would have occurred if none had been treated.[18] Although that yield might at first seem quite reasonable, the prevalence or likelihood of finding minimal and mild endometriosis also must be considered, because to realize any benefit from treatment, those with disease first have to be identified. For example, using the generally accepted 30% prevalence of endometriosis in women with unexplained infertility, the number needed to treat is 40 (12/0.30), suggesting that one would have to perform diagnostic laparoscopy for 40 women to find and treat the 12 with disease and, ultimately, to achieve 1 more

pregnancy than if none of the women had laparoscopy, which, considering the costs and potential risks of laparoscopy, is difficult to justify.

Although minimal and mild endometriosis seem to have relatively little adverse effect on fertility, more advanced stages of disease, generally characterized by ovarian endometriomas and/or significant adnexal adhesions, have greater impact. However, most women with advanced endometriosis also have signs or symptoms suggesting the disease, or endometriomas that can be imaged easily by transvaginal ultrasonography.

### *Adnexal Adhesions*

Adnexal adhesions can adversely affect fertility without causing gross tubal obstruction, but women most likely to have such adhesions have risk factors, such as previous pelvic infection or a positive CAT, pelvic pain, other evidence of advanced endometriosis, or previous pelvic or adnexal surgery, including myomectomy or ovarian cystectomy in particular.

Today, for all of the aforementioned reasons, diagnostic laparoscopy should be performed not routinely, but selectively, limited to those in whom it is most likely to yield findings that will influence the choice of treatment. Laparoscopy has a relatively high yield among women with signs, symptoms, or risk factors for peritoneal disease, and those with an abnormal HSG or transvaginal ultrasonography. In women without symptoms or risk factors having a normal HSG and ultrasonography, and perhaps also a negative CAT, laparoscopy has a very low yield and can be safely omitted.

## OVARIAN RESERVE TESTS

"Ovarian reserve" is the term used to describe the size and quality of the remaining supply of oocytes. Ovarian reserve tests provide useful diagnostic and prognostic information and are now established as an important part of the evaluation of many, if not all, infertile women.

Ovarian reserve tests are aimed at identifying women having a "diminished ovarian reserve" (DOR), implying an advanced stage of follicular depletion that may result from normal aging, premature reproductive aging, or from previous ovarian trauma (surgery, radiation, chemotherapy). Moreover, because oocyte quality generally decreases with the number of follicles remaining, a DOR also implies decreased biological fertility and reproductive lifespan. An abnormal ovarian reserve test helps to explain otherwise unexplained infertility, suggests that aggressive treatment is warranted, predicts decreased sensitivity to exogenous gonadotropin stimulation, and implies a lower likelihood for achieving a successful pregnancy with treatment, regardless of age. Conversely, a normal test result suggests that the prognosis for successful treatment is commensurate with age.

Ideally, threshold values for ovarian reserve tests should be chosen to maximize specificity for DOR (the ability of the test to correctly categorize patients as unaffected) rather than sensitivity (the ability of the test to correctly categorize patients as affected), so as to minimize the risk for a false-positive result (incorrectly categorizing a patient with a normal ovarian reserve as having DOR), which could lead to inappropriate recommendations to abandon treatment or to pursue adoption or oocyte donation. Although it would be best to also avoid futile treatment in women with unrecognized DOR, the consequences are less dire. The most important performance measures of a screening test are its PPV (the probability that a woman with a positive test truly has DOR) and NPV (the probability that a woman with a negative test truly has a normal ovarian reserve), and both vary with the prevalence of the disorder (in this case, DOR) in the test population. If the

prevalence of DOR is low, as in young women, the PPV will be low, even if sensitivity and specificity are high. Conversely, if the prevalence of DOR is high, as in older women, the PPV will be high if a highly specific threshold value is chosen. As the purpose of ovarian reserve tests is to identify women with DOR, it is most useful in women who are at high risk for DOR. When the tests are applied in a low prevalence population, many women with a normal ovarian reserve will have a false-positive result and be incorrectly categorized as having DOR.

There are three measures of ovarian reserve in common clinical use—the basal serum FSH concentration, the serum antimullerian hormone (AMH) concentration, and the "antral follicle count" (AFC) as determined by transvaginal ultrasonography. All three are useful for predicting response to gonadotropin stimulation in treatment cycles involving IUI or IVF. However, they do not reliably predict who will conceive because that relates more to the quality of the individual ooyctes that ovulate or are retrieved.

### *The Basal Serum FSH Concentration*

As rising FSH levels are one of the earliest indications of reproductive aging in women, it's logical that the serum FSH concentration might be a useful ovarian reserve test, but because FSH levels vary widely across the cycle, the serum FSH concentration is best obtained during the early follicular phase (cycle day 2 to 4) when concentrations generally are highest. With the assay systems currently in use, FSH levels $> 10$ IU/L have relatively high specificity for predicting poor response to gonadotropin stimulation, but their sensitivity is rather low.[19] Simultaneous measurement of the serum estradiol concentration provides additional information that helps in the interpretation of the basal FSH level. An early elevation in serum estradiol reflects advanced early follicular development (common in aging women), which will suppress the FSH concentration, thereby potentially masking an otherwise clearly elevated FSH level suggesting DOR. When the basal FSH is normal and the estradiol concentration is elevated ($> 60$ to 80 pg/mL), the likelihood of DOR is increased and the prognosis for achieving pregnancy is decreased. When both are high, ovarian response to gonadotropin stimulation is likely to be very poor.

### *The Serum AMH Concentration*

AMH is produced by granulosa cells of preantral and small antral follicles. Consequently, levels are gonadotropin-independent and vary little within and between cycles. The serum AMH concentration has value as a measure of ovarian reserve because the number of small antral follicles correlates with the size of the remaining follicular pool; the AMH level falls progressively throughout reproductive life and becomes undetectable near the menopause. Lower AMH levels have been associated with poor response to ovarian stimulation and low oocyte yield, embryo quality, and pregnancy rates in IVF cycles. Low AMH levels ($< 0.8$ ng/mL) have high specificity and NPV for poor response to gonadotropin stimulation and suggest that a more advanced stage of follicular depletion is a likely contributing factor in the couple's infertility.[20] The serum AMH concentration is rapidly becoming the ovarian reserve test of choice, primarily because levels vary less than FSH and the test can be performed at any time of the cycle.

### *The AFC*

The number of small antral follicles in the ovaries is proportionate to the number of primordial follicles remaining and thus provides a measure of ovarian reserve. Small antral follicles are highly responsive to FSH and can be easily imaged by transvaginal ultrasonography. The AFC is determined by counting the total

number of follicles (in both ovaries) measuring 2 to 10 mm in diameter during the early follicular phase of the cycle (cycle day 2 to 4). The AFC correlates with response to gonadotropin stimulation in IVF cycles and values $<4$ have high specificity (70% to 100%) for predicting poor response.

## *Unexplained Infertility*

Unexplained infertility is a "diagnosis" of exclusion and applies when a systematic evaluation fails to identify a cause. The diagnosis requires, at a minimum, documented evidence of ovulatory function, normal semen quality, a normal uterine cavity, and bilateral tubal patency. Women with unexplained infertility also merit ovarian reserve testing to exclude DOR. In the past, a diagnosis of unexplained infertility also required a normal PCT and endometrial dating, but no longer, because those tests have proven invalid. In the past, the diagnosis also required laparoscopy to exclude endometriosis and pelvic adhesions, but today, transvaginal ultrasonography is performed to exclude advanced endometriosis (eg, endometriomas) and laparoscopy is performed selectively, not routinely. Consequently, much of infertility that previously was attributed to cervical factors, LPD, and mild endometriosis or adhesions is now "unexplained."

There are two more potential explanations for unexplained infertility. First, there truly is no abnormality and the couple's natural fertility is at the extreme lower end of the normal range, or suffers only from the effects of reproductive aging. Second, there is a specific cause, but not one that can be identified with existing diagnostic tests, with functional abnormalities of sperm or oocyte function, fertilization, implantation or embryo development being the most logical and likely possibilities (Table 1).

## *Summary*

The modern infertility evaluation includes tests aimed at detecting the most common causes of infertility, which are anovulation, poor semen quality, abnormalities of female reproductive anatomy, and aging. In general, the most objective and useful tests are a serum progesterone concentration (if appropriate timed), a semen analysis, an HSG (or sonohysterography, depending on the importance of determining tubal patency), and a serum AMH concentration. All are easily obtained within the interval spanning a single menstrual cycle, allowing clinicians to complete a basic but still thorough evaluation quickly and easily. The PCT and endometrial histologic dating are no longer recommended and diagnostic laparoscopy should be applied selectively, in women having signs, symptoms, or imaging that suggest important and treatable pathology.

***TABLE 1. The Basic Infertility Evaluation***

| | Traditional Method | Modern Method(s) |
|---|---|---|
| Ovulation | Serum progesterone concentration | Serum progesterone concentration |
| Semen quality | Semen analysis | Semen analysis |
| Cervical factor | Postcoital test | No longer recommended |
| Uterus | Hysterosalpingography | Hysterosalpingography/sonohysterography |
| | Endometrial biopsy and dating | No longer recommended |
| Fallopian tubes | Hysterosalpingography | Hysterosalpingography |
| Peritoneum | Diagnostic laparoscopy | Selective diagnostic laparoscopy |
| Ovarian reserve | Not previously evaluated | Serum AMH concentration |

AMH indicates antimullerian hormone.

## *References*


1. Practice Committee of the American Society for Reproductive MedicineDefinitions of infertility and recurrent pregnancy loss. *Fertil Steril*. 2008; 90 (5 suppl):S60.
2. Dunson DB, Colombo B, Baird DD. Changes with age in the level and duration of fertility in the menstrual cycle. *Hum Reprod*. 2002;17:1399–1403.
3. Abraham GE, Maroulis GB, Marshall JR. Evaluation of ovulation and corpus luteum function using measurements of plasma progesterone. *Obstet Gynecol*. 1974;44:522–525.
4. Filicori M, Butler JP, Crowley WF. Neuroendocrine regulation of the corpus luteum in the human: evidence for pulsatile progesterone secretion. *J Clin Invest*. 1984;73:1638–1647.
5. Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med*. 2001;345:1388–1393.
6. Cooper TG, Noonan E, von Eckardstein S, et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*. 2010;16:231–245.
7. Kruger TF, Acosta AA, Simmons KF, et al. Predictive value of abnormal sperm morphology in in vitro fertilization. *Fertil Steril*. 1988;49: 112–117.
8. Practice Committee of the American Society for Reproductive Medicine. Report on optimal evaluation of the infertile male. *Fertil Steril*. 2006;86 (5 suppl 1):S202–S209.
9. Glazener CM, Ford WC, Hull MG. The prognostic power of the post-coital test for natural conception depends on duration of infertility. *Hum Reprod*. 2000;15:1953–1957.
10. Oei SG, Helmerhorst FM, Bloemenkamp KW, et al. Effectiveness of the postcoital test: randomised controlled trial. *BMJ*. 1998;317: 502–505.
11. Soares SR, Barbosa dos Reis MM, Camargos AF. Diagnostic accuracy of sonohysterography, transvaginal sonography, and hysterosalpingography in patients with uterine cavity diseases. *Fertil Steril*. 2000;73:406–411.
12. Noyes RW, Hertig AW, Rock J. Dating the endometrial biopsy. *Fertil Steril*. 1950;1:3–25.
13. Murray MJ, Meyer WR, Zaino RJ, et al. A critical analysis of the accuracy, reproducibility, and clinical utility of histologic endometrial dating in fertile women. *Fertil Steril*. 2004;81:1333–1343.
14. Coutifaris C, Myers ER, Guzick DS, et al. Histological dating of timed endometrial biopsy tissue is not related to fertility status. *Fertil Steril*. 2004;82:1264–1272.
15. Luttjeboer F, Harada T, Hughes E, et al. Tubal flushing for subfertility. *Cochrane Database Syst Rev*. 2007;3:CD003718.
16. Evers JLH, Land JA, Mol BW. Evidence-based medicine for diagnostic questions. *Semin Reprod Med*. 2003;21:9–15.
17. Mol BWJ, Swart P, Bossuyt PMN, et al. Is hysterosalpingography an important tool in predicting fertility outcome? *Fertil Steril*. 1997;67: 663–669.
18. Jacobson TZ, Barlow DH, Koninckx PR, et al. Laparoscopic surgery for subfertility associated with endometriosis. *Cochrane Database Syst Rev*. 2002;4:CD001398.
19. Broekmans FJ, Kwee J, Hendriks DJ, et al. A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update*. 2006;12:685–718.
20. Ficicioglu C, Kutlu T, Baglam E, et al. Early follicular antimullerian hormone as an indicator of ovarian reserve. *Fertil Steril*. 2006;85:592–596.