Testing and interpreting measures of ovarian reserve: a committee opinion

Practice Committee of the American Society for Reproductive Medicine

American Society for Reproductive Medicine, Birmingham, Alabama

Currently there is no uniformly accepted definition of decreased ovarian reserve (DOR), as the term may refer to three related but distinctly different outcomes: oocyte quality, oocyte quantity, or reproductive potential. Available evidence concerning the performance of ovarian reserve tests is limited by small sample sizes, heterogeneity among study design, analyses and outcomes, and the lack of validated outcome measures. (Fertil Steril® 2015;■:■–■. ©2015 by American Society for Reproductive Medicine.)

The process of reproductive aging centers on the generally accepted principle that human oocytes peak in number during fetal life, undergo atresia thereafter, and do not regenerate. Although female fertility declines with age, it is difficult to predict the pace of reproductive decline in an individual woman. Nonetheless, clinicians often are asked for advice on fertility potential and recommendations for fertility treatments. This document reviews the evidence relating to the clinical utility and predictive value of ovarian reserve testing. An understanding of the limitations of screening tests in general, and ovarian reserve tests in particular, is required to avoid confusion and misinterpretation, or misuse of results.

WHAT IS OVARIAN RESERVE?

Clearly, women of the same age can have very different responses to ovarian stimulation and have differing reproductive potential. The concept of ovarian reserve views reproductive potential as a function of the number and quality of remaining oocytes. Decreased or diminished ovarian reserve (DOR) describes women of reproductive age having regular menses whose response to ovarian stimulation or fecundity is reduced compared with women of comparable age. Decreased ovarian reserve is distinct from menopause or premature ovarian failure (also referred to as primary ovarian insufficiency) (1). Although ovarian reserve tests have been applied widely, debate continues over the ability of tests currently in use to predict three related, but distinctly different, outcomes: oocyte quality, oocyte quantity, and fecundity.

In most cases, the cause(s) of DOR are unknown. It is unclear whether DOR represents a pathologic condition resulting from abnormally rapid atresia in a normal pool of oocytes, from normal atresia of an abnormally small initial pool of oocytes, or simply the extreme end of a normal bell-shaped population distribution of the number of oocytes at a given age. A loss of oocytes and fertility potential are associated with exposure to systemic chemotherapy, pelvic irradiation, and genetic abnormalities (e.g., 45,X chromosomal mosaicism, FMR1 premutations). Decreased ovarian reserve has not been associated with other lifestyle behaviors, with the possible exception of cigarette smoking (2).

WHY MEASURE OVARIAN RESERVE?

Although oocyte number and quality decline with age, fertility varies significantly among women of a similar age. Consequently, a number of tests involving biochemical measures and ovarian imaging, collectively known as ovarian reserve tests, have been proposed to help predict ovarian reserve and/or reproductive potential. In women with regular menses, ovarian reserve tests do not predict whether they are entering menopause or perimenopause or distinguish whether they are experiencing a decline in

Received December 1, 2014; accepted December 2, 2014.

Reprint requests: Practice Committee of the American Society for Reproductive Medicine, 1209 Montgomery Hwy., Birmingham, Alabama 35216 (E-mail: ASRM@asrm.org).

Fertility and Sterility® Vol. ■, No. ■, ■ 2015 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2014.12.093

fertility that is pathologic or expected. When caring for a couple with infertility, clinicians use factors such as age and diagnoses to counsel individual patients and tailor the treatment plan. The goal of ovarian reserve testing is to add more prognostic information to the counseling and planning process so as to help couples choose among treatment options. However, it is important to emphasize that ovarian reserve tests are not infallible and should not be the sole criteria used to deny patients access to assisted reproductive technology (ART) or other treatments. Evidence of DOR does not necessarily equate with inability to conceive.

WHAT ARE MEASURES OF OVARIAN RESERVE?

Ovarian reserve tests include both biochemical tests and ultrasound imaging of the ovaries. Biochemical tests of ovarian reserve can be divided further into basal measurements, including measurement of follicle-stimulating hormone (FSH), estradiol, inhibin B, and antimüllerian hormone (AMH), and provocative tests such as the clomiphene citrate challenge test (CCCT). Biochemical measures of ovarian reserve are intended to probe and to reflect the biology of the aging ovary, the one component of the reproductive system most closely related to decreased fecundity.

Inhibin B and AMH are glycoprotein hormones produced by small ovarian follicles and are therefore direct measures of the follicular pool. Whereas AMH is primarily secreted by primary, preantral, and antral follicles, inhibin B is secreted primarily by preantral follicles. As the number of ovarian follicles declines with age, both AMH and early follicular phase inhibin B concentrations decline. Decreased inhibin B secretion lowers the level of central negative feedback, resulting in increased pituitary FSH secretion and in higher late luteal and early follicular FSH concentrations (an "indirect" measure). In turn, the earlier increase in FSH levels stimulates an earlier onset of new follicular growth and increase in estradiol concentrations, ultimately decreasing the length of the follicular phase and the overall cycle. Dynamic ovarian reserve tests assess the response of the hypothalamic–pituitary–ovarian axis to a stimulus.

Ultrasonographic measures of ovarian reserve include antral follicle count (AFC) and ovarian volume. The AFC describes the total number of follicles measuring 2–10 millimeters in diameter that are observed during an early follicular phase transvaginal scan. The number of antral follicles correlates with the size of the remaining follicular pool and the number of oocytes retrieved following stimulation. Ovarian volume declines with age and is therefore another potential indicator of ovarian reserve.

HOW ARE OVARIAN RESERVE TESTS USED?

Historically, ovarian reserve tests were intended to be used to screen patients before beginning a cycle of in vitro fertilization (IVF) and to treat only those patients falling within a normal range as defined by each center. However, studies examining the performance of ovarian reserve tests have used heterogeneous patient populations and outcome measures, vastly complicating their interpretation. Whereas some have screened a general IVF population, others have targeted populations of older women, seeking to discriminate women with good prognoses from those with poor prognoses despite their similar chronologic age.

Measures of ovarian reserve have been used to predict DOR, but DOR has been defined in different ways, including both reduced fecundability (the ability to achieve pregnancy) and poor ovarian response to gonadotropin stimulation. Measures of ovarian response such as the number of follicles, the number of oocytes retrieved, the number of embryos, and cancellation rate are surrogates for the clinically important outcomes: pregnancy and live birth. These surrogate outcomes are related to the clinically important outcomes but are not synonymous. Heterogeneity in study populations and varied exposures and outcomes have resulted in a wide range of test characteristics for measures of ovarian reserve reported in the literature. Therefore, the reported "effectiveness" of ovarian reserve tests as screening tests varies. Accordingly, it is important to consider study designs carefully when applying the results of these screening studies to individual patients.

BASIC PRINCIPLES OF SCREENING TESTS

The purpose of a screening test is to identify persons at risk for a disease. The purpose of using ovarian reserve testing as a screening test is to identify infertility patients at risk for DOR, who are more likely to exhibit a "poor" response to gonadotropin stimulation and to have a lesser chance of achieving pregnancy with ART, most commonly IVF.

A screening test has a number of test characteristics, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), all of which change with the diagnostic threshold, or cutpoint, used to classify an individual as being at risk for DOR (e.g., FSH $\geq$11.4 mIU/mL) (3). Good screening tests have validity. Sensitivity and specificity are two measures of test validity (Fig. 1). A valid test correctly categorizes persons who have disease as test positive (highly sensitive) and those without disease as test negative (highly specific). In other words, a highly sensitive test would capture all of the patients who have DOR. Changing cutpoints to optimize sensitivity would minimize the number of false negatives (patients with DOR categorized as normal) but increase false-positive test results (patients with normal ovarian reserve categorized as having

FIGURE 1

		Decreased ovarian reserve	
		+	-
Ovarian reserve test result	Test +	A (True positives)	B (False positives)
	Test -	C (False negatives)	D (True negatives)

2 × 2 table to calculate test characteristics of a screening test. Sensitivity = A/A+C; Specificity = D/B+D; Positive predictive value = A/A+B; Negative predictive value = D/C+D.

Practice Committee. Ovarian reserve testing. Fertil Steril 2015.

DOR). A highly specific test would correctly identify all of the patients who do not have DOR. Changing cutpoints to optimize specificity would minimize false positives but increase false negatives.

Graphically, the sensitivity and specificity of different cutpoints of a diagnostic test can be plotted as receiver operating characteristic (ROC) curves. These curves help identify the cutpoint that maximizes sensitivity and specificity. However, the diagnostic threshold that "optimally" balances specificity and sensitivity for identifying patients at risk for DOR necessarily sacrifices some specificity to improve sensitivity and thus may not be the best choice for clinical care.

For clinical application, the threshold for considering an ovarian reserve test "abnormal" should have high specificity for DOR. Specificity is the test characteristic that should be optimized to decrease false positives, or wrongly categorizing patients with normal ovarian reserve as having DOR. For the clinician, a highly specific test helps to avoid over-aggressive treatment in patients with normal ovarian reserve. Additionally, it will avoid recommending adoption or oocyte donation to patients who may have the potential to have their own genetic offspring. Because sensitivity is sacrificed when specificity is optimized, a highly specific test for DOR also would result in more women attempting IVF outcomes not knowing that the prognosis is poor.

Positive predictive value and NPV are screening test characteristics that change with the prevalence of disease (DOR) in the study population. The PPV is the probability that a woman who tests positive truly has DOR. The NPV is the probability that a woman who tests negative has normal ovarian reserve.

The most important test characteristics of a screening ovarian reserve test are its predictive values rather than sensitivity or specificity. Although predictive value is determined by sensitivity and specificity, it also is dependent on the prevalence of DOR in the population. This principle is important in determining whom to screen. If the prevalence or risk of DOR is low (e.g., in young women), the PPV (the probability that a woman who tests positive truly has DOR) will be low, even if the sensitivity and specificity are high. If the prevalence of DOR is high (e.g., in older women), the PPV will be high if a highly specific test cutpoint is chosen. Therefore, it is obvious that ovarian reserve testing is most useful in identifying DOR in women at high risk for DOR. Ideally, for tests of ovarian reserve to be clinically useful for patient counseling, the test characteristics and prevalence of DOR in a specific population or clinic should be known. The wide range of values reported in the literature makes it difficult to use these measures clinically.

The use of a screening test for DOR in a population at low risk for the condition poses several problems. Most importantly, many women will be categorized as having DOR who, in fact, have a normal ovarian reserve. The implications are important when screening women who have not yet received infertility treatment and those who may simply be curious about their reproductive potential. Ovarian reserve testing in women at low risk for DOR will yield a larger number of false-positive results (lower PPV) (4). The problem is compounded in the use of home tests, where a qualified medical professional is not readily available to interpret and explain the results.

Basal FSH

Basal serum FSH concentrations increase on day 2, 3, or 4 of the menstrual cycle with advancing reproductive age. However, assays for FSH have significant inter- and intra-cycle variability that limits their reliability (5–7). The overall correlation among different FSH assays is excellent, but absolute values can differ from one another (8). A change in the reference standard, from a human menopausal gonadotropin standard (International Reference Preparation [IRP]-hMG) to the World Health Organization (WHO) Second International Standard (IRP 78/549), complicated the generalizability of FSH cutpoints. Sample conversion of IRP-hMG values to IRP 78/549 values are as follows: high FSH 25 mIU/mL (IRP-hMG) = 16.7 (IRP 78/549), moderately high FSH 17 mIU/mL (IRP-hMG) = 11.4 (IRP 78/549), normal FSH $<$15 mIU/mL (IRP-hMG) = $<$10 mIU/mL (IRP 78/549) (8). Thus, clinicians may find it difficult to generalize FSH cutpoints reported in the medical literature to their practices unless they are using the very same assay and reference preparation (7).

Despite its limitations, FSH is commonly used as a measure of ovarian reserve, and high values have been associated with, but do not necessarily predict, both poor ovarian stimulation and the failure to conceive (8). Assays standardized against the WHO Second International Standard demonstrate high specificity (83%–100% range) for predicting poor response to stimulation (usually defined as $<$2–3 follicles or $\leq$4 retrieved oocytes) using multiple cutpoints above 10 IU/L (10–20 IU/L) (8). However, sensitivity for identifying women who will respond poorly varies widely (10%–80%) and decreases with increasing FSH cutpoints (8). Using similar cutpoints, FSH is far less sensitive for predicting the failure to achieve pregnancy. A recent study employing efficiency curves demonstrated 100% specificity for failure to achieve a live birth at FSH values above 18 IU/L (9). Cutpoints that yield high specificity (80%–100%) have low sensitivities (10%–30%) (8). Consequently, the majority of women who are tested (including those with DOR) will not have an abnormal FSH value. The test still is clinically useful, because one can be fairly certain that women having an abnormally elevated FSH value will have DOR. The PPV of FSH for poor response to ovarian stimulation or failure to conceive is higher in older women.

High FSH levels have not been associated with an increased risk of aneuploidy in pregnancies resulting from IVF (10, 11). Although FSH rises with increasing reproductive age, it remains unknown whether high FSH levels in women of reproductive age predict an earlier onset of menopause (12).

The variability in FSH levels often prompts clinicians to repeat the test. Whereas consistently elevated FSH concentrations confer a poor prognosis (13), a single elevated FSH value in women $<$40 years of age may not predict a poor response to stimulation or failure to achieve pregnancy (13). Limited evidence suggests that women with fluctuating FSH levels

should not wait for the "ideal" cycle, wherein the FSH concentration is normal, to undergo IVF stimulation (5, 14).

It has been reported that basal FSH has limited utility as a screening test (8, 15, 16). At high cutpoints that maximize specificity, sensitivity is moderate for poor response to stimulation and very low for failure to achieve pregnancy. Although relatively few women with DOR will test abnormally if cutpoints are high, those who do have an abnormal test are very likely to have DOR.

In summary, a single FSH value has very limited reliability because of inter- and intra-cycle variability (particularly if it is not elevated). An elevated FSH value has good specificity but may represent a false positive especially when used in a low-risk population. Given the inter-assay variability of FSH, the cutpoint selected by an IVF program ideally should be based on its own data or on data from studies using the same FSH assay (Table 1).

Estradiol

As a test of ovarian reserve, basal estradiol on day 2, 3, or 4 of the menstrual cycle has poor inter- and intra-cycle reliability (17). The vast majority of studies have found that basal estradiol does not differ between women with and without DOR, regardless of whether the measured outcome is poor response to ovarian stimulation or failure to achieve pregnancy (18–28). Basal estradiol alone should not be used to screen for DOR. The test has value only as an aid to correct interpretation of a "normal" basal serum FSH value. As discussed earlier, an early rise in serum estradiol concentrations is a classic characteristic of reproductive aging and can lower an otherwise elevated basal FSH level into the normal range, thereby causing a misinterpretation of the test. When the basal FSH concentration is "normal" but the estradiol level is elevated (>60–80 pg/mL) in the early follicular phase, there is limited evidence for an association with poor response, increased cancellation rates, or lower pregnancy rates (28–30).

Clomiphene Citrate Challenge Test

The CCCT involves measurements of serum FSH before (cycle day 3) and after (cycle day 10) treatment with clomiphene citrate (100 mg daily, cycle days 5–9). Whereas rising inhibin B and estradiol levels derived from a growing cohort of ovarian follicles will suppress FSH in women with responsive ovaries, the smaller follicular cohorts that can be recruited in women with DOR will generate less inhibin B and estradiol, resulting in decreased negative feedback inhibition of FSH secretion and higher stimulated FSH concentrations. An elevated FSH concentration after clomiphene stimulation therefore suggests DOR. Studies of CCCT results have observed significant inter-cycle variability in stimulated FSH levels and in the difference between basal and stimulated estradiol and inhibin B concentrations, which limits the reliability of the CCCT (6, 31, 32). A systematic review examined the ability of the CCCT to predict poor ovarian response or pregnancy after IVF over a range of day-10 FSH levels (10–22 IU/L) in women at low, average, and high risk for DOR. For the outcome of poor ovarian response, the specificity of day-10 FSH concentrations ranged between 47% and 98% and sensitivity varied between 35% and 93% (33). For the outcome of failure to achieve pregnancy, specificity has been found to range between 67% and 100% and sensitivity between 13% and 66%, depending on the study (33). In other words, out of 10 females who do not conceive through IVF, between 1 and 7 women would have had an abnormal day-10 FSH value (sensitivity) and of 10 women who do conceive, 7–10 would have a normal day-10 FSH value. In studies comparing the test performance of basal (cycle day 3) and stimulated (cycle day 10) FSH values,

TABLE 1

Summary of the value of screening tests of ovarian reserve.

		Poor response		Non-pregnancy				
Test	**Cutpoint**	**Sensitivity (%)**	**Specificity (%)**	**Sensitivity (%)**	**Specificity (%)**	**Reliability**	**Advantages**	**Limitations**
FSH	10–20 IU/L	10–80	83–100	7–58	43–100	Limited	Widespread use	Reliability Low sensitivity
AMH	0.2–0.7 ng/mL	40–97	78–92	[a]	[a]	Good	Reliability	Limit of detectability Two commercial assays Does not predict non-pregnancy
AFC	3–10	9–73	73–100	8–33	64–100	Good	Reliability Widespread use	Low sensitivity
Inhibin B	40–45 pg/mL	40–80	64–90	[a]		Limited		Reliability Does not predict non-pregnancy
CCCT (day-10 FSH)	10–22 IU/L	35–98	68–98	23–61	67–100	Limited	Higher sensitivity than basal FSH	Reliability Limited additional value to basal FSH Requires drug administration

Note: Laboratories ELISA.
[a] Insufficient evidence.

Practice Committee. Ovarian reserve testing. Fertil Steril 2015.

stimulated FSH levels have higher sensitivity but lower specificity than basal FSH concentrations (33). Compared with basal FSH and AFC, the clomiphene-stimulated day-10 FSH level does not clearly improve test accuracy for predicting poor ovarian response or pregnancy after IVF (32–34).

In summary, basal measures of FSH may be preferable to the CCCT, unless one is using the test to purposely increase sensitivity (Table 1).

Antimüllerian Hormone

Serum concentrations of AMH, produced by granulosa cells of early follicles, are gonadotropin-independent and therefore remain relatively consistent within and between menstrual cycles in both normal, young ovulating women and in women with infertility (17, 35–37).

Antimüllerian hormone was assayed previously using primarily two different assay kits (38). The current commercially available assay kit that is based on different technology has replaced the older assays (39, 40). Although the results obtained with the two earlier kits are highly correlated, the standard curves are not parallel, and there is no universally applicable conversion factor (41). Therefore, cutpoints developed and reported for one commercial AMH assay are not generalizable to the other commercial assay(s). When applying AMH cutpoints in clinical practice, clinicians must be very careful to determine that the assay used to measure AMH is the same as that used in the reference study population. Moreover, results can have a high coefficient of variation and vary among different commercial laboratories using the same assay (42).

Studies of AMH as a screening test for ovarian reserve have involved three different study populations—general IVF population, subpopulation of women at low risk for DOR, and subpopulations of women at high risk for DOR. Overall, lower AMH levels have been associated with, but do not necessarily predict, poor responses to ovarian stimulation, poor embryo quality, and poor pregnancy outcomes in IVF (43–47). Studies that correlate different mean AMH levels with IVF outcomes do not provide useful AMH cutpoints for clinical care (18, 44, 45, 48).

In various studies of general IVF populations, low AMH cutpoints (0.2–0.7 ng/mL DSL ELISA) have been found to have sensitivities ranging between 40% and 97% and specificities varying from 78% to 92% for <3 follicles or ≤ 2–4 retrieved oocytes (19, 43, 49, 50). The PPVs of these cutpoints for the same outcomes vary between 22% and 88%. The NPVs are high, between 97% and 100%, but these cutpoints are neither sensitive nor specific for predicting pregnancy (19, 49, 50). The range of test characteristics and the variable prevalence of DOR in different studies make it difficult to use these measures clinically. Ideally, site-specific data should be used to counsel patients.

Studies restricted to women at low risk for DOR were small and used exclusion criteria such as an elevated FSH, older age, anovulation, and severe male factor (51, 52). Outcomes have varied from ≤ 5 retrieved oocytes to clinical pregnancy per oocyte retrieval. Cutpoints of 2.5–2.7 ng/mL have 83% sensitivity, 82% specificity, 67%–77% PPV, and 61%–87% NPV for clinical pregnancy (20, 51). In other words, an AMH <2.7 ng/mL would correctly predict nonpregnancy in 6–8 of 10 women but would be wrong in 2–4 women (PPV). A cutpoint of 1.4 ng/mL had 76% sensitivity, 86% specificity, and 67% PPV for ≤ 5 retrieved oocytes (52). These higher AMH cutpoints decrease the specificity for DOR and, because of low prevalence of decreased ovarian reserve, resulted in lower PPV.

Several studies have restricted the sample population to women at high risk of DOR by recruiting older women, those with an elevated FSH, or those with a history of poor response (21, 44). Using undetectable AMH as a cutpoint resulted in 76% sensitivity and 88% specificity for predicting ≤ 3 follicles; the PPV was 68% and NPV was 92% (44). A higher AMH cutpoint of 1.25 ng/mL yielded 85% sensitivity, 63% specificity, 41% PPV, and 96% NPV for cycle cancellation (≤ 3 follicles), and 58% sensitivity, 75% specificity, 76% PPV, and 57% NPV for poor response (≤ 4 oocytes or cycle cancellation) (21). The limitation of applying AMH in high-risk populations is that some subjects who have "normal" IVF outcomes have low AMH values. Because normal women and those with DOR have overlapping low to undetectable AMH values, specificity cannot be optimized to 100%, reflecting the limitation of the AMH assay threshold.

In summary, AMH is a promising screening test and is likely more useful in the general IVF population or in women at high risk for DOR than in women at low risk for DOR. Low AMH cutpoints are fairly specific for poor ovarian response, but not for pregnancy. Future studies of AMH as a screening test should incorporate larger numbers of subjects in a high-risk or general IVF population. The use of AMH as a routine screening tool for DOR in a low-risk population is not recommended (Table 1).

Antral Follicle Count

Antral follicle count is the sum of antral follicles in both ovaries, as observed with transvaginal ultrasonography during the early follicular phase. Most studies have defined antral follicles as those measuring 2–10 mm in mean diameter in the greatest two-dimensional (2D) plane; some have defined antral follicles as those measuring 3–8 mm in mean diameter. Antral follicle count has good inter-cycle reliability and inter-observer reliability in experienced centers (21, 53–56). As suggested by a meta-analysis evaluating AFC as a predictor of poor ovarian response and pregnancy after IVF, a low AFC is considered to be 3–6 total antral follicles (mean of 5.2 with SD 2.11) and is associated with poor response to ovarian stimulation during IVF, but does not reliably predict failure to conceive (57).

Across general IVF study populations of patients at low and high risk for DOR, low AFC cutpoints of 3–4 total follicles (both ovaries combined) are highly specific (73%–100%) for predicting poor ovarian response (cycle cancellation, <3–4 follicles or retrieved oocytes) (21–23, 54, 57–60) but have lower sensitivity (9%–73%). The same cutpoints are moderately specific for predicting failure to conceive (64%–100%), but sensitivity is consistently low (8%–33%). The PPV and NPV of AFC for predicting poor response have varied widely in studies of general IVF subjects. The high specificity of a low AFC makes the test useful for predicting

poor ovarian response and treatment failure, but its clinical utility is limited by its low sensitivity. Inter- and intra-observer variability also may be limiting, especially in centers having less expertise or lower-quality ultrasound equipment.

In summary, the use of AFC may help to predict poor stimulation and pregnancy outcome but should not be the sole criterion for the application of ART (Table 1).

Inhibin B

Inhibin B is not a reliable measure of ovarian reserve. Inhibin B levels rise with gonadotropin-releasing hormone (GnRH) or FSH stimulation (the basis of dynamic tests of ovarian reserve) and therefore exhibit high intra-cycle variability (21, 43, 45). Inhibin B levels also vary significantly between menstrual cycles (21).

In a general IVF population, inhibin B is lower in poor responders than in women with a normal ovarian response to stimulation (42, 61). Poor response is most commonly defined as <3–5 developing follicles, resulting in IVF cycle cancellation, or as ≤4 retrieved oocytes. Cutpoints for low inhibin B vary by study (40–141 pg/mL). Low inhibin B cutpoints in the range of 40–45 pg/mL have specificities between 64% and 90% and sensitivities between 40% and 80%. The PPV of inhibin B is generally low (19%–22%) and the NPV is high (95%–97%) in general IVF populations (43, 48). In populations at high risk for DOR, PPV can be as high as 83% (21). The large majority of studies have demonstrated that inhibin B does not discriminate between pregnancy and failure to conceive (20, 21, 24, 62, 63).

In summary, the routine use of inhibin B as a measure of ovarian reserve is not recommended (Table 1).

Ovarian Volume

Ovarian volume is calculated by measuring each ovary in three planes and using the formula for the volume of an ellipsoid (D1 × D2 × D3 × 0.52 = volume). Mean ovarian volume is the average volume of both ovaries in the same individual. Ovarian volume has limited reliability as an ovarian reserve test. Some studies report clinically significant inter-cycle variability, but this observation is not consistent (4, 21, 64). When ovarian volume is acquired and stored by three-dimensional (3D) ultrasound, intra- and inter-observer variability is minimized, but specialized equipment is required (65). Overall, ovarian volume correlates with number of follicles and retrieved oocytes but not as well with pregnancy (22, 57, 66–68). In addition, studies of ovarian volume often have excluded patients with ovarian pathology, including those with polycystic ovary syndrome, endometriomas, and large cysts (69, 70). Thus, the generalizability is limited.

Several studies have demonstrated that low ovarian volume, typically <3 mL, or low mean diameter, <2 cm, predicts poor response to ovarian stimulation with high specificity (80%–90%) and a wide range of sensitivity (11%–80%) (8, 21). The reported PPV has been as low as 17% for women at low risk for DOR (23), and as high as 53% in women at high risk for DOR (21). In general, ovarian volume has been a poor predictor of pregnancy.

In summary, ovarian volume has limited value for detection of DOR. Antral follicle count is a better imaging test to screen for DOR than ovarian volume.

Ovarian Response

The tests discussed above evaluate the ovarian reserve at a point in time, whereas ovarian response represents the effect following ovarian stimulation. Poor ovarian response to ovarian stimulation usually is identified by a reduction in follicular response to maximal stimulation during IVF, resulting in a reduced number of retrieved oocytes. In order to standardize the definition of poor ovarian response, the European Society of Human Reproduction and Embryology (ESHRE) working group convened in Bologna and proposed that two of the following criteria be present to define whether a given low response to stimulation is truly poor ovarian response: [1] advanced maternal age or any other risk factor for poor ovarian response; [2] a previous poor ovarian response; and [3] an abnormal ovarian reserve test. Two episodes of poor ovarian response after maximal stimulation are sufficient to define a patient as a poor responder in the absence of advanced maternal age or an abnormal ovarian reserve test (71). In the setting of recently demonstrated repetitive good or poor ovarian response, repeated ovarian reserve testing is unnecessary.

COMBINED OVARIAN RESERVE TESTS

Because no single measure of ovarian reserve has 100% sensitivity and specificity, biochemical and imaging measures have been combined in an effort to improve test characteristics. Summarizing the validity and reliability of such combinations of ovarian reserve tests in screening for DOR is difficult because of heterogeneity in cutpoints and the choice of measures across studies (8). Combined ovarian reserve tests pose other problems because individual tests can be highly correlated. Consequently, including more than one measure in a prediction model does not improve test characteristics consistently (22, 50, 61). Moreover, using combined tests requires clinicians to obtain all of the measures in their patients, adding to the expense of screening for DOR.

Different techniques have been used to translate the statistical significance of results obtained with combined markers to clinical significance. Some have developed high-risk scoring systems (39, 41). Other studies use multivariable regression models to predict either poor response to ovarian stimulation or the number of follicles/oocytes retrieved (22, 32, 58, 72). However, complicated equations are cumbersome to apply clinically and do not provide clear cutpoints for each ovarian reserve test included. A prospective analysis of a combination of AMH, inhibin B, and 3D assessment of AFC and ovarian volume concluded that only AFC and AMH predicted poor ovarian response, and the prediction was no better than that derived from each test individually or in combination. Notably, none of the measures predicted the failure to conceive (73).

In summary, combined ovarian reserve test models do not consistently improve predictive ability over that of single ovarian reserve tests. High-risk scoring systems that combine

two or more measures may be clinically useful but require further validation.

SUMMARY

- Currently, there is no uniformly accepted definition of DOR, as the term may refer to three related but distinctly different outcomes: oocyte quality, oocyte quantity, or reproductive potential.
- Evidence of DOR does not necessarily equate with inability to conceive.
- Available evidence concerning the performance of ovarian reserve tests is limited by small sample sizes, heterogeneity among study design, analyses and outcomes, and by the lack of validated results. The design of published studies must be examined carefully before applying the results in clinical practice.
- The use of a screening test for DOR in a population at low risk for DOR will yield a larger number of false-positive results (i.e., characterizing a woman as DOR when in fact she has normal ovarian reserve).
- Overall, FSH is the most commonly used screening test for DOR, but AFC and AMH exhibit less variability and therefore are promising predictors.
- A single FSH value has very limited reliability because of inter- and intra-cycle variability.
- There is fair evidence to refute the notion that ovarian response and pregnancy rates will be improved in cycles wherein the FSH concentration is normal among women previously exhibiting abnormally elevated values.
- There is fair evidence against the use of basal estradiol concentration as a single screening test for DOR, but there is fair evidence that the basal estradiol concentration helps in the accurate interpretation of basal FSH concentrations used to screen for DOR.
- There is fair evidence to suggest that a clomiphene citrate challenge test has mildly increased sensitivity for detecting DOR compared with basal FSH concentration.
- There is mounting evidence to support the use of AMH as a screening test for poor ovarian response, but more data are needed.
- There is emerging evidence to suggest that a low AMH level (e.g., undetectable AMH) has high specificity as a screen for poor ovarian response but insufficient evidence to suggest its use to screen for failure to conceive.
- There is fair evidence to support that a low antral follicle count (<6) has moderate to high specificity as a screening test for poor ovarian response and insufficient evidence to support the use of AFC as a screening test for failure to conceive.
- There is fair evidence against the use of basal inhibin B as a screening test for DOR.
- There is fair evidence against the use of ovarian volume as a screening test for DOR.
- Poor ovarian response to maximal stimulation during IVF reflects DOR.
- There is insufficient evidence to indicate that the combined results of multiple screening tests for decreased ovarian reserve are more useful than that of each test alone.

CONCLUSIONS

- There is insufficient evidence to recommend that any ovarian reserve test now available should be used as a sole criterion for the use of ART.
- There is good evidence to support the conclusion that the number of false-positive test results will increase when screening tests for DOR are used in low-risk populations.

Acknowledgments: This report was developed under the direction of the Practice Committee of the American Society for Reproductive Medicine as a service to its members and other practicing clinicians. Although this document reflects appropriate management of a problem encountered in the practice of reproductive medicine, it is not intended to be the only approved standard of practice or to dictate an exclusive course of treatment. Other plans of management may be appropriate, taking into account the needs of the individual patient, available resources, and institutional or clinical practice limitations. The Practice Committee and the Board of Directors of the American Society for Reproductive Medicine have approved this report.

The following members of the ASRM Practice Committee participated in the development of this document. All Committee members disclosed commercial and financial relationships with manufacturers or distributors of goods or services used to treat patients. Members of the Committee who were found to have conflicts of interest based on the relationships disclosed did not participate in the discussion or development of this document.

Samantha Pfeifer, M.D.; Samantha Butts, M.D., M.S.C.E.; Daniel Dumesic, M.D.; Gregory Fossum, M.D.; Linda Giudice, M.D., Ph.D.; Clarisa Gracia, M.D., M.S.C.E.; Andrew La Barbera, Ph.D.; Randall Odem, M.D.; Margareta Pisarska, M.D.; Robert Rebar, M.D.; Richard Reindollar, M.D.; Mitchell Rosen, M.D.; Jay Sandlow, M.D.; Michael Vernon, Ph.D.; Eric Widra, M.D.

REFERENCES

1. Cooper AR, Baker VL, Sterling EW, Ryan ME, Woodruff TK, Nelson LM. The time is now for a new approach to primary ovarian insufficiency. Fertil Steril 2012;95:1890–7.
2. Cooper GS, Baird DD, Hulka BS, Weinberg CR, Savitz DA, Hughes CL Jr. Follicle-stimulating hormone concentrations in relation to active and passive smoking. Obstet Gynecol 1995;85:407–11.
3. Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update 2006;12:685–718.
4. Barnhart K, Osheroff J. Follicle stimulating hormone as a predictor of fertility. Curr Opin Obstet Gynecol 1998;10:227–32.
5. Jayaprakasan K, Campbell B, Hopkisson J, Clewes J, Johnson I, Raine-Fenning N. Establishing the intercycle variability of three-dimensional ultrasonographic predictors of ovarian reserve. Fertil Steril 2008;90:2126–32.
6. Scott RT Jr, Hofmann GE, Oehninger S, Muasher SJ. Intercycle variability of day 3 follicle-stimulating hormone levels and its effect on stimulation quality in in vitro fertilization. Fertil Steril 1990;54:297–302.
7. Kwee J, Schats R, McDonnell J, Lambalk CB, Schoemaker J. Intercycle variability of ovarian reserve tests: results of a prospective randomized study. Hum Reprod 2004;19:590–5.

8. Esposito MA, Coutifaris C, Barnhart KT. A moderately elevated day 3 FSH concentration has limited predictive value, especially in younger women. Hum Reprod 2002;17:118–23.
9. Scott RT Jr, Elkind-Hirsch KE, Styne-Gross A, Miller KA, Frattarelli JL. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. Fertil Steril 2008;89:868–78.
10. Thum MY, Abdalla HI, Taylor D. Relationship between women's age and basal follicle-stimulating hormone levels with aneuploidy risk in in vitro fertilization treatment. Fertil Steril 2008;90:315–21.
11. Massie JA, Burney RO, Milki AA, Westphal LM, Lathi RB. Basal follicle-stimulating hormone as a predictor of fetal aneuploidy. Fertil Steril 2008; 90:2351–5.
12. Soules MR, Sherman S, Parrott E, Rebar R, Santoro N, Utian W, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). Fertil Steril 2001;76:874–8.
13. Roberts JE, Spandorfer S, Fasouliotis SJ, Kashyap S, Rosenwaks Z. Taking a basal follicle-stimulating hormone history is essential before initiating in vitro fertilization. Fertil Steril 2005;83:37–41.
14. Abdalla H, Thum MY. Repeated testing of basal FSH levels has no predictive value for IVF outcome in women with elevated basal FSH. Hum Reprod 2006;21:171–4.
15. Bancsi LF, Broekmans FJ, Mol BW, Habbema JD, te Velde ER. Performance of basal follicle-stimulating hormone in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. Fertil Steril 2003;79:1091–100.
16. Jain T, Soules MR, Collins JA. Comparison of basal follicle-stimulating hormone versus the clomiphene citrate challenge test for ovarian reserve screening. Fertil Steril 2004;82:180–5.
17. Fanchin R, Taieb J, Lozano DH, Ducot B, Frydman R, Bouyer J. High reproducibility of serum anti-Mullerian hormone measurements suggests a multi-staged follicular secretion and strengthens its role in the assessment of ovarian follicular status. Hum Reprod 2005;20:923–7.
18. Hazout A, Bouchard P, Seifer DB, Aussage P, Junca AM, Cohen-Bacrie P. Serum antimullerian hormone/mullerian-inhibiting substance appears to be a more discriminatory marker of assisted reproductive technology outcome than follicle-stimulating hormone, inhibin B, or estradiol. Fertil Steril 2004;82:1323–9.
19. Ficicioglu C, Kutlu T, Baglam E, Bakacak Z. Early follicular antimullerian hormone as an indicator of ovarian reserve. Fertil Steril 2006;85:592–6.
20. Eldar-Geva T, Ben-Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, et al. Dynamic assays of inhibin B, anti-Mullerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum Reprod 2005;20:3178–83.
21. McIlveen M, Skull JD, Ledger WL. Evaluation of the utility of multiple endocrine and ultrasound measures of ovarian reserve in the prediction of cycle cancellation in a high-risk IVF population. Hum Reprod 2007;22:778–85.
22. Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD, te Velde ER. Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. Fertil Steril 2002; 77:328–36.
23. Frattarelli JL, Lauria-Costab DF, Miller BT, Bergh PA, Scott RT. Basal antral follicle number and mean ovarian diameter predict cycle cancellation and ovarian responsiveness in assisted reproductive technology cycles. Fertil Steril 2000;74:512–7.
24. Hall JE, Welt CK, Cramer DW. Inhibin A and inhibin B reflect ovarian function in assisted reproduction but are less useful at predicting outcome. Hum Reprod 1999;14:409–15.
25. Penarrubia J, Balasch J, Fabregues F, Carmona F, Casamitjana R, Moreno V, et al. Day 5 inhibin B serum concentrations as predictors of assisted reproductive technology outcome in cycles stimulated with gonadotrophin-releasing hormone agonist-gonadotrophin treatment. Hum Reprod 2000; 15:1499–504.
26. Phophong P, Ranieri DM, Khadum I, Meo F, Serhal P. Basal 17beta-estradiol did not correlate with ovarian response and in vitro fertilization treatment outcome. Fertil Steril 2000;74:1133–6.
27. Ranieri DM, Quinn F, Makhlouf A, Khadum I, Ghutmi W, McGarrigle H, et al. Simultaneous evaluation of basal follicle-stimulating hormone and 17 beta-estradiol response to gonadotropin-releasing hormone analogue stimulation: an improved predictor of ovarian reserve. Fertil Steril 1998;70:227–33.
28. Evers JL, Slaats P, Land JA, Dumoulin JC, Dunselman GA. Elevated levels of basal estradiol-17beta predict poor response in patients with normal basal levels of follicle-stimulating hormone undergoing in vitro fertilization. Fertil Steril 1998;69:1010–4.
29. Licciardi FL, Liu HC, Rosenwaks Z. Day 3 estradiol serum concentrations as prognosticators of ovarian stimulation response and pregnancy outcome in patients undergoing in vitro fertilization. Fertil Steril 1995;64:991–4.
30. Smotrich DB, Widra EA, Gindoff PR, Levy MJ, Hall JL, Stillman RJ. Prognostic value of day 3 estradiol on in vitro fertilization outcome. Fertil Steril 1995;64: 1136–40.
31. Hannoun A, Abu Musa A, Awwad J, Kaspar H, Khalil A. Clomiphene citrate challenge test: cycle to cycle variability of cycle day 10 follicle stimulating hormone level. Clin Exp Obstet Gynecol 1998;25:155–6.
32. Hendriks DJ, Broekmans FJ, Bancsi LF, de Jong FH, Looman CW, te Velde ER. Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve. Hum Reprod 2005;20:163–9.
33. Hendriks DJ, Mol BW, Bancsi LF, te Velde ER, Broekmans FJ. The clomiphene citrate challenge test for the prediction of poor ovarian response and non-pregnancy in patients undergoing in vitro fertilization: a systematic review. Fertil Steril 2006;86:807–18.
34. Yanushpolsky EH, Hurwitz S, Tikh E, Racowsky C. Predictive usefulness of cycle day 10 follicle-stimulating hormone level in a clomiphene citrate challenge test for in vitro fertilization outcome in women younger than 40 years of age. Fertil Steril 2003;80:111–5.
35. Tsepelidis S, Devreker F, Demeestere I, Flahaut A, Gervy C, Englert Y. Stable serum levels of anti-Mullerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. Hum Reprod 2007;22:1837–40.
36. La Marca A, Stabile G, Artenisio AC, Volpe A. Serum anti-Mullerian hormone throughout the human menstrual cycle. Hum Reprod 2006;21: 3103–7.
37. Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, te Velde ER, Broekmans FJ. Anti-Mullerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. J Clin Endocrinol Metab 2006;91: 4057–63.
38. La Marca A, Sighinolfi G, Radi D, Argento C, Baraldi E, Artenisio AC, et al. Anti-mullerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). Hum Reprod Update 2010;16:113–30.
39. Coulter Beckman. AMH Gen II Elisa package insert. Brea, CA: Beckman Coulter; 2011.
40. Nelson SM, La Marca A. The journey from the old to the new AMH assay: how to avoid getting lost in the values. Reprod Biomed Online 2011;4: 411–20.
41. Freour T, Mirallie S, Bach-Ngohou K, Denis M, Barriere P, Masson D. Measurement of serum anti-Mullerian hormone by Beckman Coulter ELISA and DSL ELISA: comparison and relevance in assisted reproduction technology (ART). Clin Chim Acta 2007;375:162–4.
42. Rustamov O, Smith A, Roberts SA, Yates AP, Fitzgerald C, Krishnan M, et al. Anti-Mullerian hormone: poor assay reproducibility in a large cohort of subjects suggests sample instability. Hum Reprod 2012;27:3085–91.
43. Muttukrishna S, McGarrigle H, Wakim R, Khadum I, Ranieri DM, Serhal P. Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? BJOG 2005;112: 1384–90.
44. Muttukrishna S, Suharjono H, McGarrigle H, Sathanandan M. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? BJOG 2004;111:1248–53.
45. van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bansci LF, de Jong FH, et al. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002;17:3065–71.
46. Silberstein T, MacLaughlin DT, Shai I, Trimarchi JR, Lambert-Messerlian G, Seifer DB, et al. Mullerian inhibiting substance levels at the time of HCG

administration in IVF cycles predict both ovarian reserve and embryo morphology. Hum Reprod 2006;21:159–63.
47. Ebner T, Sommergruber M, Moser M, Shebl O, Schreier-Lechner E, Tews G. Basal level of anti-Mullerian hormone is associated with oocyte quality in stimulated cycles. Hum Reprod 2006;21:2022–6.
48. Seifer DB, MacLaughlin DT, Christian BP, Feng B, Shelden RM. Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 2002; 77:468–71.
49. Gnoth C, Schuring AN, Friol K, Tigges J, Mallmann P, Godehardt E. Relevance of anti-Mullerian hormone measurement in a routine IVF program. Hum Reprod 2008;23:1359–65.
50. Penarrubia J, Fabregues F, Manau D, Creus M, Casals G, Casamitjana R, et al. Basal and stimulation day 5 anti-Mullerian hormone serum concentrations as predictors of ovarian response and pregnancy in assisted reproductive technology cycles stimulated with gonadotropin-releasing hormone agonist-gonadotropin treatment. Hum Reprod 2005;20:915–22.
51. Elgindy EA, El-Haieg DO, El-Sebaey A. Anti-Mullerian hormone: correlation of early follicular, ovulatory and midluteal levels with ovarian response and cycle outcome in intracytoplasmic sperm injection patients. Fertil Steril 2007; 89:1670–6.
52. Kwee J, Schats R, McDonnell J, Themmen A, de Jong F, Lambalk C. Evaluation of anti-Mullerian hormone as a test for the prediction of ovarian reserve. Fertil Steril 2008;90:737–43.
53. Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. Fertil Steril 2004;81:35–41.
54. Frattarelli JL, Levi AJ, Miller BT, Segars JH. A prospective assessment of the predictive value of basal antral follicles in in vitro fertilization cycles. Fertil Steril 2003;80:350–5.
55. Pache TD, Wladimiroff JW, de Jong FH, Hop WC, Fauser BC. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. Fertil Steril 1990;54:638–42.
56. Scheffer GJ, Broekmans FJ, Bancsi LF, Habbema JD, Looman CW, te Velde ER. Quantitative transvaginal two and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound Obstet Gynecol 2002;20:270–5.
57. Hendriks DJ, Mol BW, Bancsi LF, te Velde ER, Broekmans FJ. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. Fertil Steril 2005;83:291–301.
58. Chang MY, Chiang CH, Hsieh TT, Soong YK, Hsu KH. Use of the antral follicle count to predict the outcome of assisted reproductive technologies. Fertil Steril 1998;69:505–10.
59. Kupesic S, Kurjak A, Bjelos D, Vujisic S. Three-dimensional ultrasonographic ovarian measurements and in vitro fertilization outcome are related to age. Fertil Steril 2003;79:190–7.
60. Ng EH, Tang OS, Ho PC. The significance of the number of antral follicles prior to stimulation in predicting ovarian responses in an IVF programme. Hum Reprod 2000;15:1937–42.
61. Balasch J, Creus M, Fabregues F, Carmona F, Casamitjana R, Ascasco C, et al. Inhibin, follicle-stimulating hormone, and age as predictors of ovarian response in in vitro fertilization cycles stimulated with gonadotropin-releasing hormone agonist-gonadotropin treatment. Am J Obstet Gynecol 1996;175:1226–30.
62. Creus M, Penarrubia J, Fabregues F, Vidal E, Carmona F, Casamitjana R, et al. Day 3 serum inhibin B and FSH and age as predictors of assisted reproduction treatment outcome. Hum Reprod 2000;15:2341–6.
63. Smeenk JM, Sweep FC, Zielhuis GA, Kremer JA, Thomas CM, Braat DD. Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. Fertil Steril 2007;87:223–6.
64. Elter K, Sismanoglu A, Durmusoglu F. Intercycle variabilities of basal antral follicle count and ovarian volume in subfertile women and their relationship to reproductive aging: a prospective study. Gynecol Endocrinol 2005;20: 137–43.
65. Merce LT, Gomez B, Engels V, Bau S, Bajo JM. Intraobserver and interobserver reproducibility of ovarian volume, antral follicle count, and vascularity indices obtained with transvaginal 3-dimensional ultrasonography, power Doppler angiography, and the virtual organ computer-aided analysis imaging program. J Ultrasound Med 2005;24:1279–87.
66. Schild RL, Knobloch C, Dorn C, Fimmers R, van der Ven H, Hansmann M. The role of ovarian volume in an in vitro fertilization programme as assessed by 3D ultrasound. Arch Gynecol Obstet 2001;265:67–72.
67. Merce LT, Barco MJ, Bau S, Troyano JM. Prediction of ovarian response and IVF/ICSI outcome by three-dimensional ultrasonography and power Doppler angiography. Eur J Obstet Gynecol Reprod Biol 2007;132:93–100.
68. Kwee J, Elting ME, Schats R, McDonnell J, Lambalk CB. Ovarian volume and antral follicle count for the prediction of low and hyper responders with in vitro fertilization. Reprod Biol Endocrinol 2007;5:9.
69. Lass A, Skull J, McVeigh E, Margara R, Winston RM. Measurement of ovarian volume by transvaginal sonography before ovulation induction with human menopausal gonadotrophin for in-vitro fertilization can predict poor response. Hum Reprod 1997;12:294–7.
70. Jarvela IY, Sladkevicius P, Kelly S, Ojha K, Campbell S, Nargund G. Quantification of ovarian power Doppler signal with three-dimensional ultrasonography to predict response during in vitro fertilization. Obstet Gynecol 2003; 102:816–22.
71. Ferraretti AP, La Marca A, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L, et al. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. Hum Reprod 2011;26: 1616–24.
72. Lorusso F, Vicino M, Lamanna G, Trerotoli P, Serio G, Depalo R. Performance of different ovarian reserve markers for predicting the numbers of oocytes retrieved and mature oocytes. Maturitas 2007;56:429–35.
73. Jayaprakasan K, Campbell B, Hopkisson J, Johnson I, Raine-Fenning N. A prospective, comparative analysis of Antimullerian hormone, inhibin-B, and three dimensional ultrasound determinants of ovarian reserve in the prediction of poor response to controlled ovarian stimulation. Fertil Steril 2010;93:855–64.