SHORT TITLE: Ovarian reserve testing
FULL TITLE: 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

CAPSULE: Markers of ovarian reserve can be useful as predictors of oocyte yield following controlled ovarian stimulation and oocyte retrieval. However, they are poor predictors of reproductive potential independent of age.

ABSTRACT: Ovarian reserve is defined as the number of oocytes remaining in the ovary, or oocyte quantity (oocyte number). Markers of ovarian reserve include hormone levels and sonographically measured features of the ovaries. These markers can be useful as predictors of oocyte yield following controlled ovarian stimulation and oocyte retrieval. However, they are poor predictors of reproductive potential independent of age. This document replaces the document of the same name, last published in 2012.

The process of reproductive aging has traditionally centered on the principle that human oocytes peak in number during fetal life, undergo ovulation or atresia thereafter, and do not regenerate. Reproductive and ovarian senescence occurs with depletion of the number of oocytes, or ovarian reserve. Women in later stages of reproductive aging (perimenopause and menopause), when menses become irregular, have lower ovarian reserve than women in earlier reproductive stages, when menstrual cycles are usually regular (Hansen 2012). Ovarian reserve can be indirectly measured using hormone levels or ultrasound imaging of the ovaries. This document reviews the evidence relating to the clinical utility and predictive value of ovarian reserve testing as predictors of reproductive potential.

WHAT IS OVARIAN RESERVE?
Ovarian reserve is defined as the number of oocytes remaining in the ovary or oocyte quantity (oocyte number). Ovarian reserve or oocyte quantity (oocyte number) is different from oocyte quality, which relates to the potential of a fertilized oocyte to result in a live-born infant. While female infants are born with approximately 500,000 to one million oocytes, follicular atresia and ovulation result in a slow depletion of oocyte number over time and menopause subsequently ensues. Ovarian reserve correlates inversely with age, but there is considerable variation in ovarian reserve among women of the same chronologic age (Hansen 2008).

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 early-follicular-phase measurements of follicle-stimulating hormone (FSH), estradiol, or inhibin B; cycle-day independent antimüllerian hormone (AMH); and provocative tests, such as the clomiphene citrate challenge test (CCCT). Biochemical measures of
Ovarian reserve testing

Ovarian reserve are intended to directly or indirectly measure the oocyte or follicular pool.

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 early 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 an increase in estradiol concentrations, ultimately decreasing the length of the follicular phase and the overall cycle. These hormone tests are therefore used as markers of ovarian reserve.

Basal serum FSH and Estradiol

Basal serum FSH concentrations are elevated on day 2, 3, or 4 of the menstrual cycle in women with diminished ovarian reserve (DOR). Elevated basal serum FSH is a specific, but not sensitive, test for DOR (Jain 2004). However, FSH levels have significant inter- and intra-cycle variability that limits the reliability of a single measurement. High inter-cycle variability suggests more advanced DOR (Jayapraksas 2008, Kwee 2004). The overall correlation among different FSH assays is high, but absolute values can differ from one another (Esposito 2002). Basal estradiol alone should not be used to screen for DOR. The test has value only as an aid to the correct interpretation of a “normal” basal serum FSH value. 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), this may indicate ovarian dysfunction attributable to DOR (Evers 1998).

Clomiphene Citrate Challenge Test (CCCT)

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. Compared with basal FSH and ultrasonographically determined antral follicle count (AFC), the clomiphene-stimulated day-10 FSH level is not superior to non-dynamic tests for predicting poor ovarian response or pregnancy after IVF (Hendriks 2005, Hendriks 2006, Yanushpolsky 2003) nor unassisted conception in women with infertility (Haadsma 2008). For this reason, this test should be abandoned.

Antimüllerian Hormone (AMH)
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 (Fanchin 2005, Tsepeidis 2007, La Marca 2006, Hehenkamp 2006). AMH levels may be decreased in women currently using hormonal contraceptives and, therefore, should be interpreted with caution in these patients (van den Berg 2010). AMH is a more sensitive measure of ovarian reserve as compared with FSH and tends to decline before FSH rises (de Vet 2002). For this reason, AMH has largely replaced basal FSH and estradiol level testing as a biomarker of ovarian reserve. Basal FSH and estradiol levels may provide additional information in women with very low AMH levels.

Antral Follicle Count and Ovarian Volume
Ultrasonographic measures of ovarian reserve include antral follicle count (AFC) and ovarian volume. Ovarian volume declines with age and is therefore a potential indicator of ovarian reserve; however, this is uncommonly used for clinical prediction given the inter- and intra-cycle variability and lack of sensitivity (Jayaprakasan 2008). AFC 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. AFC has low inter-cycle variability and high inter-observer reliability in experienced centers (Eldar-Geva 2005, Kwee 2008, Bancsi 2004, Fratterelli 2003, Pache 1990).

SELECTING A MARKER OF OVARIAN RESERVE
AMH levels appear to be a more sensitive marker of ovarian reserve compared with early-follicular-phase hormone levels. AMH tends to decline before FSH rises (de Vet 2002), thus, high FSH is a specific marker for DOR but fails to detect a more subtle decline in ovarian reserve. While provocative testing with the CCCT may increase sensitivity for diminished ovarian reserve (Hendriks 2006), AMH testing is simpler to administer. AFC and AMH have been shown in multiple studies to be equivalent (La Marca 2014). When performed in an experienced center, AFC is a reasonable alternative to AMH.

DO OVARIAN RESERVE TESTS PREDICT REPRODUCTIVE POTENTIAL AMONG WOMEN WITH UNPROVEN FERTILITY?
Ovarian reserve declines with age, as do fertility rates. This has led to the presumption that ovarian reserve should predict reproductive potential; however, this has not been borne out in the literature. A number of studies have been conducted to assess markers of ovarian reserve as markers of current reproductive potential. In these prospective, cohort studies, markers of ovarian reserve were poor predictors of reproductive potential as measured by fecundability (the probability of conceiving in a given menstrual cycle), cumulative probability of pregnancy, or incidence of infertility.

In the EAGER (Effects of Aspirin in Gestation and Reproduction) trial, enrolled women aged 18-40 years with a history of one or more pregnancy losses (N=1,202). Women with low AMH levels (<1ng/mL) had
similar cumulative pregnancy rates as women with normal values (1.0-3.5ng/mL) (Zarek 2016). Time to Conceive, a prospective cohort study including 750 women aged 30-44 years, without a known history or risk factors for infertility, found that women with low AMH levels (<0.7ng/mL) or high FSH values (>10 IU/L) had similar cumulative pregnancy rates following 6 and 12 cycles of attempting pregnancy compared to women with normal levels (Steiner 2017 JAMA). Similarly, markers of ovarian reserve do not predict the probability of conceiving following unmedicated donor-sperm insemination cycles (Ripley 2015).

**DO OVARIAN RESERVE TESTS PREDICT REPRODUCTIVE POTENTIAL AMONG WOMEN WITH INFERTILITY?**

Based on the limited data available, it appears that results of ovarian reserve tests are no more useful than age alone in predicting unassisted pregnancy in women with infertility, nor do they offer clinically meaningful improvements over already established pregnancy prediction models, such as the Hunault model. The Hunault model incorporates female age, duration of subfertility, previous pregnancy, semen analysis results, and referral status to direct infertile patients into treatment or expectant management depending on their chances for spontaneous conception over the ensuing 12 months (van der Steeg HR 2007). Two large, prospective cohort studies found that while certain ovarian reserve tests (FSH and AFC) did predict unassisted pregnancy, they failed to substantially change patient management as compared to the Hunault model alone (van der Steeg JECM 2007, Haadsma 2008).

**DO MARKERS OF OVARIAN RESERVE IN INFERTILE WOMEN PREDICT TREATMENT SUCCESS FOLLOWING OVARIAN STIMULATION AND INTRAUTERINE INSEMINATION?**

Women with unexplained infertility are commonly treated with ovarian stimulation (OS) using oral agents or gonadotropins combined with intrauterine insemination (IUI). The goal of OS is to achieve multifollicular development to increase the odds of conception. Most studies examining the value of markers of ovarian reserve to predict success following treatment with OS/IUI have been retrospective analyses of clinical practice. Findings from these studies have been inconclusive, with some showing that DOR reduces the probability of pregnancy following OS/IUI and others showing no association (Dinelli 2014, Magendzo 2006, Tremellen 2010). Recently, a secondary analysis of the AMIGOS trial, a randomized clinical trial which enrolled 900 women with unexplained infertility and treated them with OS/IUI, found that AMH did not predict pregnancy rates following OS/IUI (Hansen 2016). In summary, the best evidence to date suggests that markers of ovarian reserve do not predict the likelihood of success following OS/IUI; however, further research is needed.

**DO OVARIAN RESERVE TESTS PREDICT OOCYTE YIELD FOLLOWING CONTROLLED OVARIAN HYPERSTIMULATION FOR IVF?**

The ability of AMH and AFC to predict oocyte yield as well as poor and excessive response to gonadotropin stimulation in assisted reproductive technologies has been well demonstrated in numerous studies (La Marca 2010, Seifer 2002, Broer 2011). Often, AMH and AFC are examined as linear variables in dose-
response models. While both continue to show excellent promise to predict ovarian response in IVF, data does not support the use of a low threshold to deter or refuse treatment. An analysis of national data assessing over 5,000 fresh autologous cycles in which the patient’s AMH was extremely low (<0.16 ng/mL) found that 54% of cycles were canceled, 50.7% of cycles resulted in 3 or fewer oocytes, 25% had no embryo to transfer; the live-birth rate was 9.5% per cycle start (20.5% per transfer) (Fleming 2015). Therefore, extremely low AMH values may be used to appropriately counsel women regarding suboptimal response and yield, but should not be used to refuse treatment (Seifer 2015).

DO OVARIAN RESERVE TESTS PREDICT PREGNANCY AND LIVE BIRTH FOLLOWING IVF?

While the ability of ovarian reserve tests, particularly AMH and AFC, to predict quantitative outcomes such as oocyte yield and ovarian responsiveness has been well shown, the same is not true for qualitative outcomes. Studies examining AMH as a prognostic measure for oocyte quality or clinical pregnancy and live birth rates are conflicting. Meta-analyses and studies with larger data sets, however, reveal similar results: AMH as an independent variable is only weakly predictive of pregnancy at best (Broer 2013 FNS, Illiodromiti 2014, Tal 2015, Tal 2018). Much of these data are in fresh transfers, yet one study investigating qualitative outcomes in frozen autologous transfers came to the same conclusion (Tal 2018). In multiple logistic regression models, AMH is clearly an important part of the puzzle but there are many more variables, such as female age, sperm quality, embryo development, stimulation protocols, and procedural and lab techniques that play a significant role.

COMBINED OVARIAN RESERVE TESTS

Attempts to improve the predictive value of ovarian reserve tests by using multiple markers instead of a single baseline marker have not been found to be beneficial (Jirge 2011, Manusson 2017, Jayaprakasan 2010, Lensen 2018). High-risk scoring systems (La Marca 2010 HRU, Nelson 2011) have been explored. Other studies have used multivariable regression models to predict either poor response to ovarian stimulation or the number of follicles/oocytes retrieved (McIlveen 2007, Hannoun 1998, Hendriks 2005, Ferraretti 2011). However, complicated equations are cumbersome to apply clinically and do not provide clear cut-points 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 the prediction was no better than that derived from each test individually (Lorusso 2007). Another study found that including AFC and AMH into predictive models for poor response improved the model fit, but that similar accuracy was reached using AMH or AFC alone (Broer 2013 HRU). Another report observed no difference in live birth when using an AFC versus AFC-plus-AMH algorithm to determine gonadotropin dosing in IVF (Magnusson 2017).

Despite a lack of evidence that more ovarian reserve testing improves outcomes, it is common for patients to have multiple screening tests performed. In addition to adding to overall expense, discordant results may complicate management and confuse the patient. Although generally positively correlated, AMH and AFC
may be discordant up to 30% of the time, according to at least one study (Li 2014 PloS). Additional research is needed to guide management in the event of discordant results.

WHAT IS POOR OVARIAN RESPONSE?

The tests discussed above reflect ovarian reserve which tends to correlate with oocyte yield, whereas ovarian response is defined as the actual oocyte yield following ovarian stimulation. Poor ovarian response 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, a 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 (≥ 40 years of 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 (Jarvela 2003). In the setting of recently demonstrated repetitive good or poor ovarian response, additional ovarian reserve testing is unnecessary.

WHICH PATIENTS SHOULD UNDERGO OVARIAN RESERVE TESTING?

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 prognostic information to the counseling and planning process to help couples choose among treatment options as ovarian reserve testing is only testing quantity, not quality, of the remaining oocyte pool. 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 or other treatments (Scott 2008). Evidence of DOR does not necessarily equate with inability to conceive.

Markers of ovarian reserve should not be used as a “fertility test” for women who are not infertile or who have untested fertility. Consequently, markers of ovarian reserve should not be used to promote planned oocyte cryopreservation. Decisions regarding oocyte cryopreservation should be based on a woman’s reproductive plans and her age. Age is a much stronger predictor of reproductive success than is ovarian reserve. All healthcare providers are encouraged to speak with their patients about their reproductive plans at all stages of reproductive life. Opening this dialogue and educating patients facilitates their decision making.
SUMMARY

- Dynamic tests such as the CCCT do not improve test accuracy for predicting poor ovarian response, pregnancy after IVF, or unassisted conception over basal markers and, therefore, should be abandoned.
- Currently, AMH and AFC are the most sensitive and reliable markers of ovarian reserve.
- Combined ovarian reserve test models do not consistently improve predictive ability over that of single ovarian reserve tests.
- Markers of ovarian reserve do not predict current reproductive potential among women with unproven fertility.
- Results of ovarian reserve tests are not useful in predicting the likelihood of unassisted pregnancy in women with infertility, nor do they offer clinically meaningful improvements over already established pregnancy prediction models.
- Markers of ovarian reserve do not appear to predict pregnancy following ovarian stimulation with intrauterine insemination for unexplained infertility.
- The ability of AMH and AFC to predict oocyte yield as well as poor and excessive ovarian responsiveness in IVF has been well demonstrated.
- Extremely low AMH values should not be used to refuse treatment in IVF.
- AMH and AFC have only a weak association with qualitative outcomes such as oocyte quality, clinical pregnancy rates, and live birth rates.
- Poor ovarian response to maximal stimulation during IVF reflects DOR and further ovarian reserve testing is unnecessary.

CONCLUSIONS

AMH and AFC are currently the simplest, most sensitive, and specific measures of ovarian reserve. Markers of ovarian reserve have been shown to be good predictors of oocyte yield, however, poor independent predictors of reproductive potential. Therefore, they should not be used as a “fertility test” or to deny access to infertility treatment.

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.

This document was reviewed by ASRM members and their input was considered in the preparation of the final document. The Practice Committee acknowledges the special contribution of Amber Cooper, MD, MSCI and Violet Klenov, MD in the preparation of this document. 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.

References


6. 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.


