Human Reproduction Vol.21, No.11 pp. 2729–2735, 2006 doi:10.1093/humrep/del188
Advance Access publication August 26, 2006.

DEBATE—CONTINUED

Assessment of ovarian reserve—should we perform tests of ovarian reserve routinely?

Abha Maheshwari[1], Paul Fowler and Siladitya Bhattacharya

Department of Obstetrics and Gynecology, Aberdeen University, Assisted Conception Unit, Aberdeen Maternity Hospital, Aberdeen, UK

[1]To whom correspondence should be addressed at: Department of Obstetrics and Gynecology, Aberdeen University, Assisted Conception Unit, Aberdeen Maternity Hospital, Aberdeen AB25 2ZL, UK. E-mail: abha.maheshwari@abdn.ac.uk

Women undergoing IVF are routinely subjected to one or more tests of ovarian reserve. The results of these tests are also being extrapolated to women attending infertility clinics and those planning to delay childbearing. This debate examines the predictive power of currently available tests of ovarian reserve and questions the value of subjecting women to ovarian reserve tests. We propose that in the absence of an agreement on (i) a definition of poor ovarian reserve, (ii) the population to be tested and (iii) which interventions are effective in women with poor ovarian reserve, routine ovarian reserve testing is unhelpful.

Key words: assisted reproduction/FSH/infertility/oocyte quality/ovarian reserve

Introduction

Ovarian reserve is a term used to describe the functional potential of the ovary and reflects the number and quality of oocytes within it. In future, such a test may help in counselling women concerned about the loss of reproductive potential due to other conditions such as treatment for cancer (Macklon and Fauser, 2005).

Virtually, the entire literature on ovarian reserve focuses on the ability of the ovary to respond to gonadotrophin stimulation, in the context of IVF. A good test of ovarian reserve should be predictive of conception (with or without treatment) and should indicate how long current levels of ovarian activity can be maintained before ovarian ageing sets in. In a subfertile population attending for fertility treatment, a test of ovarian reserve should guide us in prognosticating outcome in individual cases by (i) predicting the chances of pregnancy and live birth with or without treatment and (ii) selecting an optimal dose of ovarian stimulation where treatment using ovarian stimulation is planned. As such, an effective test of ovarian reserve could have a pivotal role in guiding management throughout the fertility workup and treatment by facilitating individualized clinical decision making. We believe that it is time to critically appraise the quality and feasibility of the available tests of ovarian reserve in terms of their desired purpose in (i) IVF treatment and (ii) non-IVF subfertility population and (iii) general.

Tests of ovarian reserve

Over the years, various tests and markers of ovarian reserve have been described in the literature (Table I). We searched the literature using the key word ‘ovarian reserve’ (Medline 1966–2006). A total of 257 articles were found. Further searches for individual tests of ovarian reserve were made using key words individually (FSH, Follicle stimulating hormone, AMH, anti-Mullerian hormone, inhibin-B, oestradiol, AFC, antral follicle count, ovarian volume, CCCT, clomiphene citrate challenge test, GAST, GnRH agonist stimulation test, EFORT, exogenous FSH ovarian reserve test and ovarian biopsy). Appropriate cross-references were manually searched as well.

FSH

Most IVF units use basal FSH levels (measured on day 3 of the menstrual cycle) as an indicator of ovarian responsiveness, even though the evidence to support its efficacy as a routine test is weak (Wolff and Taylor, 2004). Early-follicular-phase fluctuations in FSH are a reflection of the balance between ovarian steroid and peptide inhibition and the hypothalamo-pituitary drive during the period just before the selection of the dominant follicle. Day 3 FSH is an indirect measure of the size of the follicle cohort and is regulated by various factors, including inhibins, activins, estradiol and follistatins (te Velde and Pearson, 2002). From a pathophysiological point of view, large inter-cycle variations in basal FSH remain a frequent problem. Appropriate timing of FSH measurement is difficult for women with irregular periods, such as those with polycystic ovary syndrome (PCOS). Despite appropriately timed methods of sample collection, inter-cycle variations and inter-sample variations (within assay and between assays) may result in disparate FSH measurements (Lambalk and de Koning, 1998). A wide range (4–25 IU) in threshold values have been

Table I. Markers of ovarian reserve

Static tests
Age
Basal serum FSH
Basal serum estradiol
Basal LH/FSH ratio
Basal serum inhibin-B level
Basal serum anti-Müllerian hormone level
Basal ovarian volume
Basal antral follicle count
Ovarian stromal blood flow
Ovarian biopsy
Dynamic tests
Clomiphene citrate challenge test (CCCT)
GnRH agonist stimulation test (GAST)
Exogenous FSH ovarian reserve test (EFORT)

used for abnormal levels of basal FSH (Bancsi *et al*., 2003). The limitation of FSH in estimating ovarian reserve and counselling patients has been recognized (Sharara *et al*., 1998), and the usefulness of FSH as a routine test in the prediction of IVF outcome has been questioned before (Bancsi *et al*., 2003). There is some evidence to support the predictive value of FSH in a population of women at high risk (women >40 years of age, women with poor response to ovarian stimulation and women who have failed to conceive in previous cycles) in terms of the likelihood of achieving pregnancy through assisted reproduction (Barnhart and Osheroff, 1999). In contrast, the role of day 3 FSH in the evaluation of young healthy women is extremely limited (Wolff and Taylor, 2004).

Discrepancies also exist between the predictive value of FSH in terms of ovarian response to gonadotrophin stimulation as opposed to the likelihood of an ongoing pregnancy (Bancsi *et al*., 2003), reflecting the limitation of using a test of ovarian responsiveness as a marker of ovarian reserve. A meta-analysis of studies on basal FSH as a predictor of pregnancy has confirmed a wide range of sensitivities (0.03–0.85), specificities (0.20–1.00) and likelihood ratios (0.5–13.4) (Bancsi *et al*., 2003). Predictions with a substantial change from pre-FSH-test probability to post-FSH-test probability are only achieved at extreme threshold levels of basal FSH and thus applicable to no more than a small minority of patients (Bancsi *et al*., 2003). Basal FSH is simple to perform but does not diagnose poor ovarian reserve until high thresholds are used. Combined with other markers, such as age and antral follicle count (AFC), FSH can be useful for counselling regarding poor ovarian response. As a test, it does not predict pregnancy and should not be used to exclude people from assisted reproduction technology (ART), especially regularly cycling young women.

AFC

Recently, AFC, as visualized by transvaginal ultrasound scan, has attracted considerable interest as a test of ovarian reserve. An age-related decline in the AFC has been observed (Ruess *et al*., 1996; Ng *et al*., 2003; van Rooij *et al*., 2005). A systematic review has demonstrated the superiority of AFC over basal FSH in the prediction of poor ovarian response (Hendriks *et al*., 2005a). Although AFC is the single best available predictor of response to ovarian stimulation with exogenous gonadotrophins, the precise definition of what constitutes an antral follicle is variable, with cited diameters ranging between 2–10 and 2–5 mm (Frattarelli *et al*., 2003; Hansen *et al*., 2003; Bancsi *et al*., 2004; Durmusoglu *et al*., 2004). Moreover, different thresholds for defining low AFC are used in different studies (Hendriks *et al*., 2005a). Inter-cycle variability has been investigated in women with proven fertility (Scheffer *et al*., 1999), those undergoing IVF (Hansen *et al*., 2003; Bancsi *et al*., 2004) and in general subfertile women (Elter *et al*., 2005). Inter-cycle variability appears to be more significant in young women and in women with high AFC. Hence, a low AFC in young, infertile but ovulatory women should be interpreted cautiously, as this may not indicate poor ovarian reserve.

AFC has been suggested to be a better marker than age and FSH for distinguishing between older patients with good and poor pregnancy prospects (Klinkert *et al*., 2005a). The sensitivity, specificity and likelihood ratio of AFC to live birth have not been tested in the literature. Data from a single study suggest that poor response to stimulation can be predicted with a sensitivity of 0.89, a specificity of 0.39 and a positive likelihood ratio of 1.45 (Muttukrishna *et al*., 2005).

The performance of AFC for predicting failure to achieve pregnancy is poor (Bancsi *et al*., 2002; Ng *et al*., 2003; Hendriks *et al*., 2005a). This is because while AFC determines the number of oocytes, a clinically relevant outcome (pregnancy or live birth) depends on oocyte quality as well as quantity.

Serum estradiol

Elevated basal estradiol may predict the poor response even when basal FSH is normal (Evers *et al*., 1998). In regularly menstruating women between the ages of 24 and 50 years, no differences in basal estradiol levels have been demonstrated according to age (Lee *et al*., 1988). No relationship has been found between serum estradiol levels and pregnancy rates (Scott *et al*., 1989). As a test for the prediction of pregnancy, basal estradiol has a likelihood ratio of 1.2–3.1 (Licciardi *et al*., 1995; Smotrich *et al*., 1995) although different thresholds are used in the two published studies. The value of cycle day 3 estradiol levels in the prediction of ovarian reserve is still debatable (Bukulmez and Arici, 2004). No data are available on the relationship between day 3 estradiol values and fecundity in spontaneous cycles.

Ovarian volume

In women with small ovaries (<3 cm^3), the cancellation rate of IVF is higher (Sharara and McClamrock, 1999). Low ovarian volume has also been found to correlate with the number of growing follicles, but not with the number of oocytes retrieved (Tomas *et al*., 1997). A correlation was found between ovarian volume and reproductive success in ART cycles; however, the likelihood ratio of a positive test with regard to pregnancy was 1.0–1.4, suggesting that its value is limited (Syrop *et al*., 1995; Lass *et al*., 1997). Moreover, there is a wide range in the definition of normal ovarian volume in the reproductive age group.

Ovarian biopsy

Ovarian biopsy (Lambalk *et al*., 2004) has not been found to be a useful routine test of ovarian reserve. Apart from being

invasive and posing unknown future adverse effects, ovarian biopsy is not a reliable test to assess reproductive ageing on fertility, as there is a highly varied distribution of the follicles throughout the ovary. The use of ovarian biopsy in predicting pregnancy has not been tested.

Inhibin-B

Inhibin-B is mainly produced by the granulosa cells in growing follicles and offers a more immediate assessment of ovarian activity than other serum tests. A fall in day 3 inhibin-B levels may predict poor ovarian reserve before the expected rise in day 3 FSH (Danforth *et al*., 1998; Seifer *et al*., 1999; Fried *et al*., 2003). However, other studies do not support its use as a predictive marker in IVF (Hall *et al*., 1999; Creus *et al*., 2000). Inhibin-B levels are influenced by the amount of fat in an individual (Tinkanen *et al*., 2001), suggesting that the follicles of obese women do not produce as much inhibin-B as those of lean women. The highest sensitivity (81%) and specificity (81%) were obtained at a serum level of 56 pg/ml where the end-point was the number of oocytes collected (Ficicioglu *et al*., 2003). Using 40 pg/ml as the threshold for being low ovarian reserve, it yielded the following values: sensitivity (87%), specificity (49%) and a positive likelihood ratio of 1.7 (Muttukrishna *et al*., 2005). The odds ratio for a clinical pregnancy (basal serum inhibin >45 versus <45 pg/ml) was 6.8 (CI 1.8–25.6) (Seifer *et al*., 1997).

Anti-Müllerian hormone

Anti-Müllerian hormone (AMH) is produced by the granulosa cells of the recruited follicles until they become sensitive to FSH (te Velde and Pearson, 2002). AMH has been identified as a regulator of the recruitment, preventing the depletion of all primordial follicle pool at once (Themmen, 2005). It has been found to decline with advancing female age (de Vet *et al*., 2002) and been suggested as a predictor of ovarian response (van Rooij *et al*., 2002; Seifer *et al*., 2002; Fanchin *et al*., 2003). Moreover, AMH is the only marker of ovarian reserve that can be tested in follicular as well as luteal phase, although the threshold levels in both phases need to be standardized.

AMH levels have been found to be two to three times higher in PCOS women (Laven *et al*., 2004; Mulders *et al*., 2004; Piltonen *et al*., 2005), making it difficult to find a threshold value for poor ovarian reserve without a significant overlap with normal values.

In a recent study (n = 56), AMH was found to be predicting pregnancy better than AFC and inhibin-B (PPV 67% for serum AMH >18 pmol/l and 39% for serum AMH <18 pmol/l) (Eldar-Geva *et al*., 2005). However, in a prospective randomized study (n = 75) despite been shown to be the most sensitive and specific indicator of ovarian response (thresholds 25 pg/l), when compared with other available tests, it did not predict pregnancy (Ficicioglu *et al*., 2006).

AMH may provide a useful marker of ovarian reserve in future although more work is needed before it can be routinely used.

Dynamic tests

Another approach towards identifying ovarian reserve involves dynamic testing. This involves taking a baseline serum sample, stimulating the ovaries (FSH/Clomiphene/GnRH agonist) and then retesting the serum level again for the same marker. All the dynamic tests are more expensive, invasive and associated with the side effects of administered stimulation regimens.

Clomiphene citrate challenge test (CCCT) involves the administration of 100 mg clomiphene citrate on days 5–9 and measurement of serum FSH on days 3 and 10. An abnormal test is defined as an abnormally high FSH on day 3 and/or on day 10. Recent meta-analysis (Jain *et al*., 2004) has shown that the CCCT is no better than basal FSH in predicting a clinical pregnancy.

Exogenous FSH ovarian reserve test (EFORT) involves the measurement of basal FSH, estradiol and estradiol response 24 h after a 300 IU FSH injection on day 3. The addition of the dynamic component to the day 3 FSH concentration might be an improvement of the predictive value of good response to ovarian stimulation (Fanchin *et al*., 1994). EFORT has not been studied for prediction of pregnancy in an IVF population. There have been no studies on EFORT as a test of ovarian reserve in general subfertile population.

GnRH agonist stimulation test (GAST) evaluates the estradiol serum concentration change from cycle day 2 to day 3 after the administration of a supraphysiological dose of a GnRH agonist. A prompt estradiol response may be associated with better ovarian reserve. Earlier ART studies did not show any significant benefit in the prediction of ovarian response (Padilla *et al*., 1990; Winslow *et al*., 1991); however, later studies did (Ranieri *et al*., 1998; Hendriks *et al*., 2005b). Although, when compared with the predictive accuracy and clinical value of the day 3 AFC and inhibin-B measurement, GAST did not perform better (Hendriks *et al*., 2005b). In addition, its predictive ability towards ongoing pregnancy is poor (Hendriks *et al*., 2005b).

CCCT has been tested in ART as well as in a general infertility population, but GAST and EFORT have not been tested outside the ART population (Bukman and Heineman, 2001). Hence, the results cannot be extrapolated to predict the fertility potential of the general population.

Combination of tests

Wolff and Taylor (2004) have suggested using a triple screen test as a model for the clinical application of day 3 FSH with or without the inclusion of other markers of ovarian reserve. Kline *et al*. (2005) produced predictive models based on chronological age, ovarian volume, FSH and inhibin-B. In fecund young women who want to defer childbearing, expanded models do not improve on the knowledge of the woman's age alone for predicting whether or not she would encounter problems when later trying to conceive. For an older woman who wants to know how long she can postpone childbearing or who is trying to conceive and wants to know whether to expect difficulties, the best model would have a positive predictive value (PPV) of 79% as opposed to a PPV of 60% based on age alone. These equations were however based on a single sample and require validation (Kline *et al*., 2005). Moreover, the equation applies to women of demonstrated fertility and needs further testing to determine whether they are useful to women seeking treatment for infertility.

Combinations of various markers (AFC, AMH and inhibin-B) have been tried, and a joint scoring system has been developed which predicts a poor response to gonadotrophin stimulation at best with 87% sensitivity and 80% specificity and a positive likelihood ratio of 4.36%. However, they have not been tested for prediction of pregnancy (Muttukrishna *et al*., 2005). There are commercially available kits for testing ovarian reserve currently available which use a combination of FSH, AMH and inhibin-B. Further work, exploring their predictive value, in the population tested, using appropriate endpoints (such as live birth) is awaited.

Interventions in poor ovarian reserve

There are few effective interventions for women who are expected to be poor responders. Oocyte cytoplasmic transfer from healthy donors to the oocyte of women with diminished ovarian reserve (FSH > 15 IU) has been tried without any significant improvement in the success rate (n = 15/18) (Opsahl *et al*., 2002). Data from small studies suggest that doubling the starting dose of gonadotrophins in those who are expected to be poor responders (on the basis of a low AFC) does not improve pregnancy rates (n = 26/26) (Klinkert *et al*., 2005b).

In a retrospective analysis of IVF cycles, Jurema *et al*. (2003) speculated that IVF outcome may be improved by testing the basal hormonal profile repeatedly in consecutive cycles and starting IVF in the cycle in which low FSH is detected, using a GnRH antagonist (Jurema *et al*., 2003). The effectiveness of this policy remains to be proven as repeated measurements of basal FSH are of no clinical value (Abdalla and Thum, 2006).

Alternative approaches have been described for IVF treatment in women with decreased ovarian reserve and include microdose GnRH agonist flare protocol or other flare (Schoolcraft *et al*., 1997; Surrey *et al*., 1998; Surrey and Schoolcraft, 2000), use of GnRH antagonists with gonadotrophins, low dose GnRH agonist suppression before gonadotrophin stimulation, assisted hatching of embryos and use of estrogen or oral contraceptives in the cycle before gonadotrophin stimulation. Unfortunately, there are no randomized trials to compare the relative efficacy of these approaches (Practice Committee of the American Society for Reproductive Medicine, 2004).

Ovarian reserve—screening, diagnostic or prognostic test?

The term test refers to any method for obtaining additional information on a patient's health status. There are different kinds of tests: (i) screening tests, (ii) diagnostic tests and (iii) prognostic tests.

The attributes of a useful screening test are summarized in Table II (Wilson and Jungner, 1968). Poor ovarian reserve does not fulfil the criteria of a disease for which a screening programme can be developed. We are still some distance away from a reliable test with high sensitivity and specificity. We still lack effective and acceptable treatments that can be used to correct poor ovarian reserve. We are uncertain about the population to be screened (whether this should include all women, women with infertility or women attending for assisted reproduction).

Table II. Criterion for a useful screening test (Wilson and Jungner, 1968)

Diagnostic test
Sensitive and specific
Safe and acceptable
Simple and cheap
Reliable
Disease
Serious
High prevalence of preclinical stage
Natural history understood
Long period between first signs and overt disease
Diagnosis and treatment
Facilities are adequate
Effective, acceptable and safe treatment available

Source: Wilson JM and Jungner YG (1968) Epidemiology and prevention. In Beaglehole R (ed.) *Basic Epidemiology*. World Health Organization, Geneva, 1993 (Reprint 2002), p. 93.

The sensitivity and PPV of a test for any condition are known to increase with its prevalence within the population (Barnhart and Osheroff, 1999). Thus, the predictive value of any test of ovarian reserve will be low in young women compared with that in those who are older and present with infertility. Currently available investigations do not appear to hold any more promise as diagnostic test. The ultimate objective of a diagnostic test is to enable the clinician to choose an adequate management strategy. Important features of a diagnostic test, as it pertains to testing ovarian reserve before infertility treatment, have been described by Jain *et al*. (2004) and are listed in Table III. The accuracy of such a test refers to the level of agreement between the information from the test under evaluation (index test) and a reference standard (Bossuyt *et al*., 2004) and may be influenced by subject characteristics such as age and co-morbidity (Knottnerus and Muris, 2003). In subfertility, we are dealing with a wide age range of 25–45 years. Furthermore, other contributing factors, such as endometriosis, increased body mass index and male factor, can confound the accuracy of the test. Unfortunately, in the case of ovarian reserve, there is no reference standard. Outcomes in the literature include (i) response to ovarian stimulation as seen on ultrasound scan, (ii) dose of gonadotrophins used, (iii) numbers of oocytes collected during IVF and (iv) pregnancy rates or (v) live birth rates. The first three are surrogate outcomes, whereas the last two are susceptible to treatment bias. There are difficulties in comparing individual studies because of differences in the definitions of poor ovarian reserve, population studied and different stimulation regimens used. These factors also limit the usefulness of these tests to prognosticate outcome (such as the likelihood of live birth) in women presenting with subfertility.

The availability of numerous candidate tests is testament to the fact that there is no single reliable test. Although a screening test would identify those women who are more likely to go on to have an early ovarian failure, a diagnostic test would identify women who currently have a poor ovarian reserve. A prognostic test would predict which of the women diagnosed with poor ovarian reserve should be excluded from treatment based on poor ovarian reserve.

From Table IV, it is clear that none of the tests fulfil the criteria for a good screening test as opportunistic screening or to

Table III. Important features of a diagnostic test, as it pertains to testing ovarian reserve (Jain *et al*., 2004)

Feature	Question addressed
Sensitivity	How good is this test in picking up women who cannot conceive?
Specificity	How good is this test at correctly excluding women who can conceive?
Positive predictive value	If a woman has an abnormal test, what is the probability that she cannot conceive?
Negative predictive value	If a woman has a normal test, what is the probability that she can conceive?
Likelihood ratio of a positive test	How much more likely is an abnormal test to be found in a woman who cannot conceive than in a woman who can?
Likelihood ratio of a negative test	How much more likely is a normal test to be found in a woman who can conceive than in a woman who cannot?

Source: Jain T, Soules MR and Collins JA (2004) Comparison of basal follicle-stimulating hormone versus the clomiphene citrate challenge test for ovarian reserve screening. Fertil Steril 82,180–185.

Table IV. Summary of tests of ovarian reserve

Test	Good screening test	Good diagnostic test	Good prognostic test
Basal serum FSH	x	?	x
Basal serum antral follicle count	x	?	x
Basal serum estradiol	x	?	x
Basal serum anti-Müllerian hormone level	Under evaluation	Under evaluation	x
Basal serum inhibin-B	x	?	x
Ovarian volume	x	?	x
Ovarian biopsy	x	?	x
Ovarian stromal blood flow	x	?	x
Clomiphene citrate challenge test (CCCT)	x	?	x
GnRH agonist stimulation test (GAST)	x	?	x
Exogenous FSH ovarian reserve test (EFORT)	x	?	x

develop a mass screening programme. Most of them will diagnose poor ovarian reserve, but only at the extreme ends of a range of values. However, those extreme values are not standardized in the literature. None of the available tests can predict who is going to get pregnant in either group (good ovarian reserve or poor ovarian reserve). Some women may experience ovarian ageing earlier than chronological age. Although it can be argued that making a clear diagnosis in these women may allow early access to ART, none of the currently available markers of ovarian ageing are sufficiently accurate to provide a sound basis for eligibility for ART (Baird *et al*., 2005).

Conclusion

Available tests for ovarian reserve do not have enough predictive power to justify their routine clinical use. All the tests of ovarian reserve described so far test oocyte quantity. Unfortunately, it has not been possible to do a direct assessment of oocyte quality (Broekmans *et al*., 1998). Most of the literature on ovarian reserve is based on the ART population and cannot be directly extrapolated to either all infertile women or the general female population of reproductive age. Even if we were able to diagnose or predict low ovarian reserve, we would need to question the use of these tests in the absence of an effective intervention (including increase in the starting dose of gonadotrophins) to improve reproductive outcome (Bukulmez and Arici, 2004). None of the available tests, or combination of tests, of ovarian reserve have been shown to predict pregnancy or live birth with sufficient accuracy. The literature emphasizes the fact that regularly cycling women should not be excluded from IVF on the basis of abnormal results following tests of ovarian reserve. This makes one question the rationale for indiscriminate testing for ovarian reserve. One of the arguments supporting these tests is that they allow us to counsel women regarding poor prognosis. The reality is that even in women with normal ovarian reserve, the chances of IVF success can be considered to be relatively low (live birth rate 20–30%). At the same time, there are other factors, including age, parity, previous treatment and the quality of treatment, which can influence pregnancy and live birth rates. In the assisted conception population, the first cycle of IVF still remains the most informative test in terms of how a woman will respond to ovarian stimulation. We may need to accept the fact that it is futile to try to identify a suitable test of ovarian reserve until we agree on its definition and a have a clear idea about its natural history.

References

Abdalla H and Thum MY (2006) Repeated testing of basal FSH levels has no predictive value of IVF outcome in women with elevated basal FSH. Hum Reprod 21,171–174.

Baird DT, Collins J, Egozcue J, Evers LH, Gianaroli L, Leridon H, Sunde A, Templeton A, Van Steirteghem A, Cohen J *et al*. ESHRE Capri Workshop (2005) Fertility and ageing. Hum Reprod Update 11,261–276.

Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD and te Velde ER (2002) Predictors of poor ovarian response in in-vitro fertilization: a prospective study comparing basal markers of ovarian reserve. Fertil Steril 77,328–336.

Bancsi LF, Broekmans FJ, Mol BW, Habbema JD and te Velde ER (2003) 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 79,1091–1100.

Bancsi LF, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2004) Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. Fertil Steril 81,35–41.

Barnhart K and Osheroff J (1999) We are overinterpreting the predictive value of serum follicle-stimulating hormone levels. Fertil Steril 72,8–9.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC *et al*. (2004) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Fam Pract 21,4–10.

Broekmans FJ, Scheffer GJ, Bancsi LF, Dorland M, Blankenstein MA and te Velde ER (1998) Ovarian reserve tests in infertility practice and normal fertile women. Maturitas 30,205–214.

Bukman A and Heineman MJ (2001) Ovarian reserve testing and the use of prognostic models in patients with subfertility. Hum Reprod Update 7,581–590.

Bukulmez O and Arici A (2004) Assessment of ovarian reserve. Curr Opin Obstet Gynecol 16,231–237.

Creus M, Penarrubia J, Fabregues F, Vidal E, Carmona F, Casamitjana R, Vanrell JA and Balasch J (2000) Day 3 serum inhibin B and FSH and age as predictors of assisted reproduction treatment outcome. Hum Reprod 15,2341–2346.

Danforth DR, Arbogast LK, Mroueh J, Kim MH, Kennard EA, Seifer DB and Friedman CI (1998) Dimeric inhibin: a direct marker of ovarian aging. Fertil Steril 70,119–123.

Durmusoglu F, Elter K, Yoruk P and Erenus M (2004) Combining cycle day 7 follicle count with the basal antral follicle count improves the prediction of ovarian response. Fertil Steril 81,1073–1078.

Eldar-Geva T, Ben-Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, Gal M, Zylber-Haran E and Margalioth EJ (2005) Dynamic assays of inhibin B, anti-Mullerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum Reprod 20,3178–3183.

Elter K, Sismanoglu A and Durmusoglu F (2005) Intercycle variabilities of basal antral follicle count and ovarian volume in subfertile women and their relationship to reproductive aging: a prospective study. Gynecol Endocrinol 20,137–143.

Evers JL, Slaats P, Land JA, Dumoulin JC and Dunselman GA (1998) 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 69,1010–1014.

Fanchin R, de Ziegler D, Olivenness F, Taieb J, Dzik A and Frydman R (1994) Exogeneous follicle stimulating hormone ovarian reserve test (EFORT): a simple and reliable screening test for detecting poor responders in in-vitro fertilization. Hum Reprod 9,1607–1611.

Fanchin R, Schonauer LM, Righini C, Guibourdenche J, Frydman R and Taieb J (2003) Serum anti-Mullerian hormone is more strongly related to ovarian follicular status than serum inhibin B, oestradiol, FSH and LH on day 3. Hum Reprod 18,323–327.

Ficicioglu C, Kutlu T, Demirbasoglu S and Mulayim B (2003) The role of inhibin B as a basal determinant of ovarian reserve. Gynecol Endocrinol 17,287–293.

Ficicioglu C, Kutlu T, Baglam E and Bakacak Z (2006) Early follicular antimullerian hormone as an indicator of ovarian reserve. Fertil Steril 85,592–596.

Frattarelli JL, Levi AJ, Miller BT and Segars JH (2003) A prospective assessment of the predictive value of basal antral follicles in vitro fertilization cycles. Fertil Steril 80,350–355.

Fried G, Remaeus K, Harlin J, Krog E, Csemiczky G, Aanesen A and Tally M (2003) Inhibin B predicts oocyte number and the ratio IGF-I/IGFBP-1 may indicate oocyte quality during ovarian hyperstimulation for in vitro fertilization. J Assist Reprod Genet 20,167–176.

Hall JE, Welt CK and Cramer DW (1999) Inhibin A and inhibin B reflect ovarian function in assisted reproduction but are less useful at predicting outcome. Hum Reprod 14,409–415.

Hansen KR, Morris JL, Thyer AC and Soules MR (2003) Reproductive aging and variability in the ovarian antral follicle count: application in the clinical setting. Fertil Steril 80,577–583.

Hendriks DJ, Mol BW, Bancsi LF, te Velde ER and Broekmans FJ (2005a) 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 83,291–301.

Hendriks DJ, Broekmans FJ, Bancsi LF, Looman CW, de Jong FH and te Velde ER (2005b) Single and repeated GnRH agonist stimulation tests compared with basal markers of ovarian reserve in the prediction of outcome in IVF. J Assist Reprod Genet 22,65–73.

Jurema MW, Bracero NJ and Garcia JE (2003) Fine tuning cycle day 3 hormonal assessment of ovarian reserve improves in vitro fertilization outcome in gonadotropin-releasing hormone antagonist cycles. Fertil Steril 80,1156–1161.

Kline J, Kinney A, Kelly A, Reuss ML and Levin B (2005) Predictors of antral follicle count during the reproductive years. Hum Reprod 20,2179–2189.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2005a) The antral follicle count is a better marker than basal follicle-stimulating hormone for the selection of older patients with acceptable pregnancy prospects after in vitro fertilization. Fertil Steril 83,811–814.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2005b) Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. Hum Reprod 20,611–615.

Knottnerus JA and Muris JW (2003) Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 56,1118–1128.

Lambalk CB and de Koning CH (1998) Interpretation of elevated FSH in the regular menstrual cycle. Maturitas 30,215–220.

Lambalk CB, de Koning CH, Flett A, Van Kasteren Y, Gosden R and Homburg R (2004) Assessment of ovarian reserve. Ovarian biopsy is not a valid method for the prediction of ovarian reserve. Hum Reprod 19,1055–1059.

Lass A, Skull J, McVeigh E, Margara R and Winston RM (1997) Measurement of ovarian volume by transvaginal sonography before ovulation induction with human menopausal gonadotrophin for in-vitro fertilization can predict poor response. Hum Reprod 12,294–297.

Laven JS, Mulders AG, Visser JA, Themmen AP, De Jong FH and Fauser BC (2004) Anti-Mullerian hormone serum concentrations in normoovulatory and anovulatory women of reproductive age. J Clin Endocrinol Metab 89,318–323.

Lee SJ, Lenton EA, Sexton L and Cooke ID (1988) The effect of age on the cyclical patterns of plasma LH, FSH, oestradiol and progesterone in women with regular menstrual cycles. Hum Reprod 3,851–855.

Licciardi FL, Liu HC and Rosenwaks Z (1995) Day 3 estradiol serum concentrations as prognosticators of ovarian stimulation response and pregnancy outcome in patients undergoing in vitro fertilization. Fertil Steril 64,991–994.

Macklon NS and Fauser BC (2005) Ovarian reserve. Semin Reprod Med 23,248–256.

Mulders AG, Laven JS, Eijkemans MJ, de Jong FH, Themmen AP and Fauser BC (2004) Changes in anti-Mullerian hormone serum concentrations over time suggest delayed ovarian ageing in normogonadotrophic anovulatory infertility. Hum Reprod 19,2036–2042.

Muttukrishna S, McGarrigle H, Wakim R, Khadum I, Ranieri DM and Serhal P (2005) Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology. BJOG 112,1384–1390.

Ng EH, Yeung WS, Fong DY and Ho PC (2003) Effects of age on hormonal and ultrasound markers of ovarian reserve in Chinese women with proven fertility. Hum Reprod 18,2169–2174.

Opsahl MS, Thorsell LP, Geltinger ME, Iwaszko MA, Blauer KL and Sherins RJ (2002) Donor oocyte cytoplasmic transfer did not enhance implantation of embryos of women with poor ovarian reserve. J Assist Reprod Genet 19,113–117.

Padilla SL, Bayati J and Garcia JE (1990) Prognostic value of the early serum estradiol response to leuprolide acetate in in vitro fertilization. Fertil Steril 53,288–294.

Piltonen T, Morin-Papunen L, Koivunen R, Perheentupa A, Ruokonen A and Tapanainen JS (2005) Serum anti-Mullerian hormone levels remain high until late reproductive age and decrease during metformin therapy in women with polycystic ovary syndrome. Hum Reprod 20,1820–1826.

Practice Committee of the American Society for Reproductive Medicine (2004) Aging and infertility in women. Fertil Steril 82,S102–S106.

Ranieri DM, Quinn F, Makhlouf A, Khadum I, Ghutmi W, McGarrigle H, Davies M and Serhal P (1998) 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 70,227–233.

van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH and Themmen AP (2002) Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 17,3065–3071.

van Rooij IA, Broekmans FJ, Scheffer GJ, Looman CW, Habbema JD, de Jong FH, Fauser BJ, Themmen AP and te Velde ER (2005) Serum antimullerian hormone levels best reflect the reproductive decline with age in normal women with proven fertility: a longitudinal study. Fertil Steril 83,979–987.

Ruess ML, Kline J, Santos R, Levin B and Timor-Tritsch I (1996) Age and the ovarian follicle pool assessed with transvaginal ultrasonography. Am J Obstet Gynecol 174,624–627.

Scheffer GJ, Broekmans FJ, Dorland M, Habbema JD, Looman CW and Te Velde ER (1999) Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. Fertil Steril 72,845–851.

Schoolcraft W, Schlenker T, Gee M, Stevens J and Wagley L (1997) Improved controlled ovarian hyperstimulation in poor responder in vitro fertilization patients with a microdose follicle-stimulating hormone flare, growth hormone protocol. Fertil Steril 67,93–97.

Scott RT, Toner JP, Muasher SJ, Oehninger S, Robinson S and Rosenwaks Z (1989) Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. Fertil Steril 51,651–654.

Seifer DB, Lambert-Messerlian G, Hogan JW, Gardiner AC, Blazar AS and Berk CA (1997) Day 3 serum inhibin-B is predictive of assisted reproductive technologies outcome. Fertil Steril 67,110–114.

Seifer DB, Scott RT, Bergh PA, Abrogast LK, Friedman CI, Mack CK and Danforth DR (1999) Women with declining ovarian reserve may demonstrate a decrease in day 3 serum inhibin B before a rise in day 3 follicle stimulating hormone. Fertil Steril 72,63–65.

Seifer DB, MacLaughlin DT, Christian BP, Feng B and Shelden RM (2002) Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 77,468–471.

Sharara FI and McClamrock HD (1999) The effect of aging on ovarian volume measurements in infertile women. Obstet Gynecol 94,57–60.

Sharara FI, Scott RT and Seifer DB (1998) The detection of diminished ovarian reserve in infertile women. Am J Obstet Gynecol 179,804–812.

Smotrich DB, Widra EA, Gindoff PR, Levy MJ, Hall JL and Stillman RJ (1995) Prognostic value of day 3 estradiol on in vitro fertilization outcome. Fertil Steril 64,1136–1140.

Surrey ES and Schoolcraft WB (2000) Evaluating strategies for improving ovarian response of the poor responder undergoing assisted reproductive techniques. Fertil Steril 73,667–676.

Surrey ES, Bower J, Hill DM, Ramsey J and Surrey MW (1998) Clinical and endocrine effects of a microdose GnRH agonist flare regimen administered to poor responders who are undergoing in vitro fertilization. Fertil Steril 69,419–424.

Syrop CH, Willhoite A and Van Voorhis BJ (1995) Ovarian volume: a novel outcome predictor for assisted reproduction. Fertil Steril 64,1167–1171.

Themmen APN (2005) Anti-Mullerian hormone: its role in follicular growth initiation and survival and as an ovarian reserve marker (2005). J Natl Cancer Inst 34,18–21.

Tinkanen H, Blauer M, Laippala P, Tuohimaa P and Kujansuu E (2001) Correlation between serum inhibin B and other indicators of the ovarian function. Eur J Obstet Gynecol Reprod Biol 94,109–113.

Tomas C, Nuojua-Huttunen S and Martikainen H (1997) Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. Hum Reprod 12,220–223.

te Velde ER and Pearson PL (2002) The variability of female reproductive ageing. Hum Reprod Update 8,141–154.

de Vet A, Laven JS, de Jong FH, Themmen AP and Fauser BC (2002) Antimullerian hormone serum levels: a putative marker for ovarian aging. Fertil Steril 77,357–362.

Winslow KL, Toner JP, Brzyski RG, Oehninger SC, Acosta AA and Muasher SJ (1991) The gonadotropin-releasing hormone agonist stimulation test – a sensitive predictor of performance in the flare-up in vitro fertilization cycle. Fertil Steril 56,711–717.

Wolff EF and Taylor HS (2004) Value of the day 3 follicle-stimulating hormone measurement. Fertil Steril 81,1486–1488.