Human Reproduction Update, Vol.17, No.1 pp. 46–54, 2011
Advanced Access publication on July 28, 2010 doi:10.1093/humupd/dmq034

human reproduction update

AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis

S.L. Broer [1,*], M. Dólleman [1], B.C. Opmeer [2], B.C. Fauser [1], B.W. Mol [3], and F.J.M. Broekmans [1]

[1]Department of Reproductive Medicine and Gynecology, University Medical Center, room F05.126, P.O. Box 85500, 3508 GA Utrecht, The Netherlands [2]Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands [3]Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam, The Netherlands

*Correspondence address. E-mail: s.l.broer@umcutrecht.nl

Submitted on February 18, 2010; resubmitted on June 22, 2010; accepted on June 30, 2010

TABLE OF CONTENTS

BACKGROUND: Anti-Mullerian hormone (AMH) is a marker of ovarian reserve status and represents a good predictor of ovarian response to ovarian hyperstimulation. The aim of this study was to assess the accuracy of AMH and antral follicle count (AFC) as predictors of an excessive response in IVF/ICSI treatment.

METHODS: A systematic review and meta-analysis of the existing literature was performed. Studies were included if 2 × 2 tables for the outcome excessive response in IVF patients in relation to AMH/AFC could be constructed. Using a bivariate meta-analytic model, both summary point estimates for sensitivity and specificity were calculated, as well as summary ROC curves. Clinical value was analysed by calculating post-test probabilities of excessive response at optimal cut-off levels, as well as the corresponding abnormal test rates.

RESULTS: Nine studies reporting on AMH and five reporting on AFC were found. Summary estimates of sensitivity and specificity for AMH were 82 and 76%, respectively, and 82 and 80%, respectively, for AFC. Comparison of the summary estimates and ROC curves for AMH and AFC showed no statistical difference. Abnormal test rates for AMH and AFC amounted to ~14 and 16%, respectively, at cut-off levels where test performance is optimal [likelihood ratio for a positive result (LR +)>8], with a post-test probability of ± 70%.

CONCLUSIONS: Both AMH and AFC are accurate predictors of excessive response to ovarian hyperstimulation. Moreover, both tests appear to have clinical value. This opens ways to explore the potential of individualized FSH dose regimens based on ovarian reserve testing.

Key words: Anti-müllerian hormone / antral follicle count / IVF / ovarian response / meta-analysis

Background

In *in vitro* fertilization (IVF) treatment, excessive response to FSH stimulation introduces the risk of abdominal discomfort, painful follicle aspirations and cycle cancellations (Delvigne and Rozenberg, 2002). An excessive response will also typically introduce the risk of ovarian hyperstimulation syndrome (OHSS), a potentially life-threatening condition (Fauser *et al*., 2008). Excessive response to ovarian stimulation will generate many oocytes for the laboratory that will not unequivocally lead to a full range of good quality embryos (Baart *et al*., 2007; Heijnen *et al*., 2007; Verberg *et al*., 2009). In addition, chances of pregnancy may decrease (van der Gaast *et al*., 2006). In view of these drawbacks, elimination of exaggerated ovarian response in stimulation protocols will improve safety, success and costs of assisted reproduction technology (ART) programs.

For primary preventive management to be developed, the reliability of tools for prediction of ovarian response needs to be assessed first. Ovarian response prediction is mainly based on ovarian reserve tests like the antral follicle count (AFC) and Anti-Mullerian hormone (AMH) (Broekmans *et al*., 2006; Broer *et al*., 2009). The AFC comprises the number of 2–5 or 2–10 mm diameter follicles measured in the ovaries at the start of the menstrual cycle (de Carvalho *et al*., 2008) and is highly correlated to the number of oocytes retrieved at pick up (Kwee *et al*., 2007; Broer *et al*., 2009). AMH has been implicated as the most valuable marker of ovarian reserve as serum concentrations correlate highly with baseline AFC and the number of oocytes retrieved at aspiration (van Rooij *et al*., 2002; Eldar-Geva *et al*., 2005; Tremellen *et al*., 2005; Nakhuda *et al*., 2006, 2007; Riggs *et al*., 2008; Nardo *et al*., 2009). The aim of the present systematic literature review was to asses the true accuracy of AMH and AFC as prognostic indicators for the prediction of an excessive response after IVF/ICSI treatment.

Methods

Search and selection strategy

The literature was searched for studies that addressed the capacity of AFC or AMH as prognostic indicators of excessive ovarian response after controlled ovarian hyperstimulation in an IVF or ICSI treatment. No pre-set definition of excessive ovarian response was used. Excessive ovarian response definition included oocytes at retrieval above a certain threshold, estrogen-level above a certain threshold, the development of OHSS or cycle cancellation due to a high response, or combinations of these. Also, any cut-off or set of cut-offs for an abnormal AMH or AFC were included in this review.

A systematic search in Medline was carried out using the keywords '*in vitro* fertilization', '*in vitro* fertilization', 'assisted', 'intracytoplasmic', ' intracytoplasmatic' in combination with 'anti-Mullerian hormone', 'Mullerian inhibiting factor', 'Mullerian inhibiting substance' or 'antral follicle count'. A period of all the years through November 2009 was covered by the search. The abstracts of all studies identified were read by one researcher (M.D.). Any article that could possibly be of value for the association between AMH and AFC and the IVF outcome excessive ovarian response was preselected. In the next step, two researchers (M.D. and S.L.B.) carefully read and judged all preselected articles independently. If it was judged possible to construct 2 × 2 tables from the data presented in the paper, the study was selected for final inclusion and analysis in this review. In a 2 × 2 or contingency table, the true positive, true negative, false positive and false negative test results at a certain cut-off are displayed. In the event of any disagreement between the two authors, the opinion of a third researcher (F.J.M.B.) was final.

The authors of studies that reported on the ovarian reserve test result in relation to IVF outcome without the possibility of constructing 2 × 2 tables were contacted by email and asked to provide the necessary data for the construction of such a table. If adequate data were obtained in this way, the study was added to the selection. In every selected study, the reference list was scanned to identify studies that could possibly be included in the selection and these were then processed as described.

Each selected study was further scored by the researchers (M.D and S.L.B.) regarding the following study quality characteristics: (i) patient sampling (consecutive or other); (ii) data collection method (prospective or retrospective); (iii) study design (cohort or case–control); (iv) blinding (present or absent); (v) selection bias, i.e. exclusion of cases based on criteria that affect the ability to generalize the findings of the study, for instance women with elevated basal FSH or women over 38 years of age (present or absent); (vi) verification bias, i.e. the use of results of the test under study in adapting the treatment protocol in order to prevent the predicted outcome, for instance poor ovarian response (present or absent); (vii) analysis upon one or multiple cycles per couple; and (viii) stimulation protocol (GnRH agonist or GnRH antagonist). Also, data on the cut-off levels used were recorded, as was the assay used for AMH measurement and whether AFC was measured in 2–5 or 2–10 mm follicles.

Because this review used only published data from the literature, no approval from our institutional review board was required.

Analysis

First, 2 × 2 tables were constructed from which sensitivity and specificity were calculated. Sensitivity-specificity points were displayed in the receiver operating characteristics (ROC) space (1-specificity versus sensitivity). Combinations of sensitivity and 1-specificity are indicative of the test accuracy, where studies reporting high accuracy for both sensitivity and specificity are located in the upper-left corner of the ROC space, and poor test results are located close to the $x = y$ line.

A meta-analysis was performed using a bivariate regression model (Reitsma *et al*., 2005). In short, this bivariate model preserves the two-dimensional nature of prognostic data in a single model, rather than using a single outcome measure for each study such as the diagnostic odds ratio. The bivariate model simultaneously estimates sensitivity and specificity, and incorporates the negative correlation that may exist between sensitivity and specificity within studies, owing to possible implicit differences in the applied threshold between studies. When necessary, the bivariate model uses a random approach for both sensitivity and specificity, allowing for heterogeneity beyond chance due to clinical or

methodological differences between studies. In addition, the model acknowledges the difference in precision by which sensitivity and specificity have been measured in each study. This means that studies with a larger number of women with an excessive response received more weight in the calculation of the pooled estimate of sensitivity, whereas studies with a high number of women without an excessive response were more influential in the pooling of specificity.

Sensitivity was plotted against 1-specificity (false positive rate) and pooled estimates for sensitivity and specificity were calculated and also plotted, together with the 95% confidence interval (CI) ellipse. As different studies have reported results for different thresholds to define a positive test (cut-off), we did not limit our analysis to a single threshold value, but took advantage of the fact that the model incorporates opposite effects on sensitivity and specificity when using different cut-offs. In order to account for dependent observations (observations on different cut-offs from the same study are likely to be correlated), we estimated the model in 250 stratified bootstrap samples, in which only one threshold value from each study was randomly selected. The overall estimates of sensitivity and specificity were based on the average from 250 bootstrap samples.

The results of the model were used to estimate summary ROC curves, where the increase in sensitivity and decrease in specificity reflect the shift in threshold value of the ovarian reserve test in the model. We thereby had to convert parameter estimates from the bivariate model to those in the Summary ROC model, as these are basically different statistical approaches for the same underlying model (Harbord *et al.*, 2007). The difference between AMH and AFC in pooled sensitivity and specificity was tested by fitting the bivariate model on data for both tests, with the test included as a covariate in the model.

To asses the clinical value of both tests, post-test probabilities for the prediction of an excessive response were calculated, by using the estimated summary ROC curve and assuming an arbitrary prevalence (or pre-test probability) of 20% for an excessive response. A series of likelihood ratio ranges for an abnormal test result was then derived from several points of the estimated summary ROC curve, and at these various ranges of likelihood ratios, the post-test probabilities for both tests were computed, as well as the corresponding abnormal test rates.

All statistical analyses were done using SAS 9.1 for Windows (Proc NlMixed in the bivariate model).

Results

Systematic review

The systematic Medline search produced 170 hits. Of these, 126 articles were excluded on the basis of title and abstract. Another 30 studies were excluded on the basis of the fully read article. Finally, 14 studies were selected to be appropriate for the current meta-analysis. From those 14 studies, in four studies 2 × 2 tables could be constructed from the article itself. The remaining 10 authors were contacted and asked for the necessary data.

Three studies could not be included as the authors did not reply to the email request (Seifer *et al.*, 2002; Tremellen *et al.*, 2005; Lekamge *et al.*, 2007). Seven authors did respond with the appropriate data to construct 2 × 2 tables. Thus, a final number of 11 studies could be included for data extraction and meta-analysis (Ng *et al.*, 2000; van Rooij *et al.*, 2002; Eldar-Geva *et al.*, 2005; Ebner *et al.*, 2006; Kwee *et al.*, 2008; La Marca *et al.*, 2007; Nelson *et al.*, 2007; Lee *et al.*, 2008; Riggs *et al.*, 2008; Nardo *et al.*, 2009; Aflatoonian *et al.*, 2009). Six studies reported on the capacity of AMH to predict excessive response after IVF, two studies reported on the capacity of AFC and three studies studied both AMH and AFC (Fig. 1).

The characteristics of the included studies are listed in Table I. From this table, it becomes clear that all studies but one presented data for one cycle per couple and that the majority used a

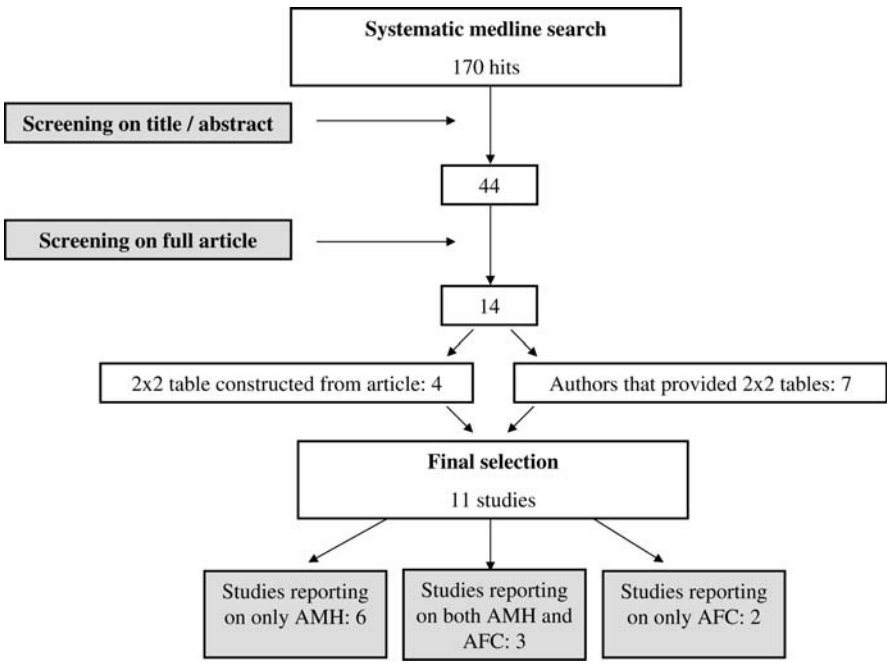

Figure 1 Search and selection strategy.

Table I Characteristics of the included studies.

Author	Test	Consecutive	Cohort /case–control	Pro-/retrospective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle	AMH assay /AFC count	Definition of excessive response
Nelson *et al.*	AMH	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	DSL	≥21 oocytes
La Marca *et al.*	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC	>16 oocytes
Ebner *et al.*	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC	≥15 oocytes
Lee *et al.*	AMH	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	DSL	OHSS
Riggs *et al.*	AMH	Yes	Cohort	retrospective	Yes	Yes	Yes	No	Yes	DSL	≥15 oocytes
Nardo *et al.*	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	DSL	≥20 oocytes, E2 > 17 000 pmol/l
Aflatoonian *et al.*	AMH & AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	BC, AFC 2–6 mm	≥15 oocytes, E2 > 3000 pg/ml
Van Rooij *et al.*	AMH & AFC	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC, AFC 2–5 mm	≥14 oocytes
Eldar Geva *et al.*	AMH & AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	BC, AFC 2–10 mm	≥14 oocytes
Kwee *et al.*	AFC	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	AFC 2–10 mm	>20 oocytes
Ng *et al.*	AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	NS	≥15 oocytes

DSL, Diagnostic System Laboratories; BC, Beckman-Coulter; NS, not stated.

prospective cohort design. However, selection bias was judged to be present in all studies. This concerned the exclusion of older women or women with signs of decreased ovarian reserve, or exclusion of cases with the PCO syndrome. The definition of excessive response was not uniform. It ranged from number of oocytes retrieved over 14 up to over 21 or the development of OHSS.

Accuracy of AMH in excessive response prediction

Sensitivities and specificities for the prediction of excessive ovarian response, as calculated from each study reporting on AMH, are summarized in Table II. A plot of sensitivity–specificity combinations in an ROC space is shown in Fig. 2. For AMH, the sensitivity varied between 40 and 95% and the specificity varied between 31 and 96%. Using the bivariate model that accounts for the heterogeneity of the studies, the summary estimates for sensitivity and specificity were calculated. The summary estimates were 82% (95% CI 52–95%) for sensitivity and 76% (95% CI 43–93%) for specificity.

Figure 2 shows the summary estimate for the overall test accuracy as calculated from the bivariate model and its 95% CI ellipse, as well as the summary ROC curve.

Accuracy of AFC in excessive response prediction

Sensitivities and specificities for the prediction of an excessive ovarian response, as calculated from each study reporting on AFC are summarized in Table II. A plot of sensitivity–specificity combinations in an ROC space is shown in Fig. 2. For the AFC, the sensitivity varied between 20 and 94% and specificity varied between 33 and 98%. Using the bivariate model that accounts for the heterogeneity of the studies, the summary estimates for sensitivity and specificity were calculated. The summary estimate of sensitivity was 82% (95% CI 30–98%) and the summary estimate of specificity was 80% (95% CI 31–97%).

Figure 2 shows the summary estimates as calculated by the bivariate model and its 95% CI ellipse, as well as the summary ROC curve for the AFC in the prediction of an excessive response.

Clinical value

On the basis of the summary ROC curves depicted in Fig. 2, a range of positive likelihood ratios was calculated corresponding to various sensitivity–specificity points on this ROC curves. For each of these likelihood ratio values, the pre AMH or AFC test probabilities of an excessive response were converted into post-test probabilities of an excessive response. Table III depicts a series of likelihood ratios ranges and the probability of obtaining an abnormal test result for AMH or AFC corresponding to this likelihood ratio range, as well as the post-test probability of an excessive response. At a positive likelihood ratio of at least ~8, the post-test probability of having an excessive response is close to 70%, if the pre-test probability is assumed to be ~20%. The probability of obtaining a test result for AMH or AFC with a likelihood ratio of at least ~8 is 14 and 16%, respectively.

Table II Performance of AMH and AFC in the prediction of excessive response.

Author	Cycles (n)	Cut-off (ng/ml)	Abnormal test result (%)	Sensitivity	Specificity	LR+	Pre-test probability	Post-test probability
AMH								
van Rooij *et al.* (2002)	114	3.50	0.08	0.40	0.95	8.32	0.09	0.44
Eldar-Geva *et al.* (2005)	53	3.50	0.32	0.72	0.89	6.32	0.34	0.76
Ebner *et al.* (2006)	135	1.66	0.75	0.95	0.31	1.38	0.16	0.21
	135	4.52	0.25	0.55	0.81	2.80	0.16	0.35
La Marca *et al.* (2007)	48	2.60	0.50	0.86	0.56	1.95	0.15	0.25
	48	7.00	0.23	0.57	0.83	3.35	0.15	0.36
Nelson *et al.* (2007)	314	2.10	0.27	0.88	0.79	4.10	0.08	0.26
	314	3.50	0.08	0.57	0.96	13.80	0.07	0.52
Lee *et al.* (2008)	262	1.99	0.50	0.90	0.62	2.38	0.23	0.42
	262	3.36	0.25	0.62	0.87	4.64	0.23	0.58
Riggs *et al.* (2008)	123	1.59	0.49	0.84	0.67	2.56	0.31	0.53
Nardo *et al.* (2009)	165	3.50	0.36	0.88	0.70	2.90	0.10	0.24
Aflatoonian *et al.* (2009)	159	4.83	0.42	0.93	0.78	4.26	0.28	0.63
AFC								
Ng *et al.* (2000)	128	9	0.39	0.60	0.71	2.09	0.31	0.49
	128	14	0.10	0.20	0.94	3.48	0.31	0.62
van Rooij *et al.* (2002)	114	14	0.42	0.92	0.63	2.49	0.10	0.22
Eldar-Geva *et al.* (2005)	56	14	0.78	0.94	0.33	1.41	0.40	0.48
Kwee *et al.* (2008)	110	10	0.38	0.94	0.71	3.26	0.15	0.36
	110	12	0.30	0.88	0.80	4.33	0.15	0.42
	110	14	0.21	0.81	0.89	7.64	0.15	0.57
	110	16	0.11	0.50	0.96	11.75	0.15	0.67
	110	18	0.06	0.31	0.98	14.69	0.15	0.71
Aflatoonian *et al.* (2009)	159	16	0.31	0.89	0.92	11.26	0.28	0.82

Note: If a study reported on multiple cut-off values, data for all cut-off values are shown. LR+, likelihood ratio for a positive test result.

Comparison of AMH with AFC

Comparison of summary point estimates for accuracy of the prediction of excessive response showed no statistically significant difference in the performance for AMH compared with the AFC, when sensitivity ($P = 0.87$) and specificity ($P = 0.80$) at the estimated summary cut-off point were considered. In the comparison of the estimated summary ROC curves, AFC seemed to perform slightly better than AMH, although the curves did not differ statistically. It should be noted that the summary curve for the AFC was based on fewer studies (Fig. 2).

Clinical value as outlined in Table III indicated a similar performance for AMH compared with the AFC. This is in line with the course of the ROC curves along the *y*-axis suggesting that many cases of an excessive response can be identified with only a limited number of false positives. For both AMH and AFC, sensitivity can amount up to 70% with a false positive rate of 15%, and this performance level will imply a realistic number of abnormal tests (~25%).

Discussion

Main findings

The current meta-analysis summarizes the available evidence concerning the accuracy of AMH and the AFC in the prediction of excessive ovarian response to stimulation for IVF. It appears that both tests have a good discriminatory capacity to separate normal and excessive responders, with a definition that varies across the studies from more than 14–21 oocytes yielded. From the ROC curves (Fig. 2) it becomes clear that, currently, AMH and AFC have an equal level of accuracy in the prediction of an excessive response and that there is no statistical difference between both tests. Moreover, both AMH and AFC have clinical value, with an abnormal test rate of 14 and 16%, respectively, at cut-off levels where test performance is optimal (LR+ >8). At these cut-off levels the post-test probability of an excessive response appeared to be close to 70%.

The comparison between the AFC and AMH for their use as predictive tests for ovarian response may imply other factors than accuracy alone. For AMH as a laboratory test, measurement stability will be dealt with according to routine procedures, but routine assays may not yet be easily available. To date, two commercially available immunoassays, the Beckmann–Coulter and DSL ELISA, exist. These assays have demonstrated a very good correlation, making it possible to translate results from one to the other within the same dataset. However, there is no obvious match in absolute levels between studies; with the Beckmann–Coulter measurements reported as being approximately four to five times higher than the DSL measurements (Bersinger *et al.*, 2007; Freour *et al.*, 2007). Therefore, standardization of these assays is urgently needed. This situation also hampers the efforts to extract a generally applicable cut-off level for

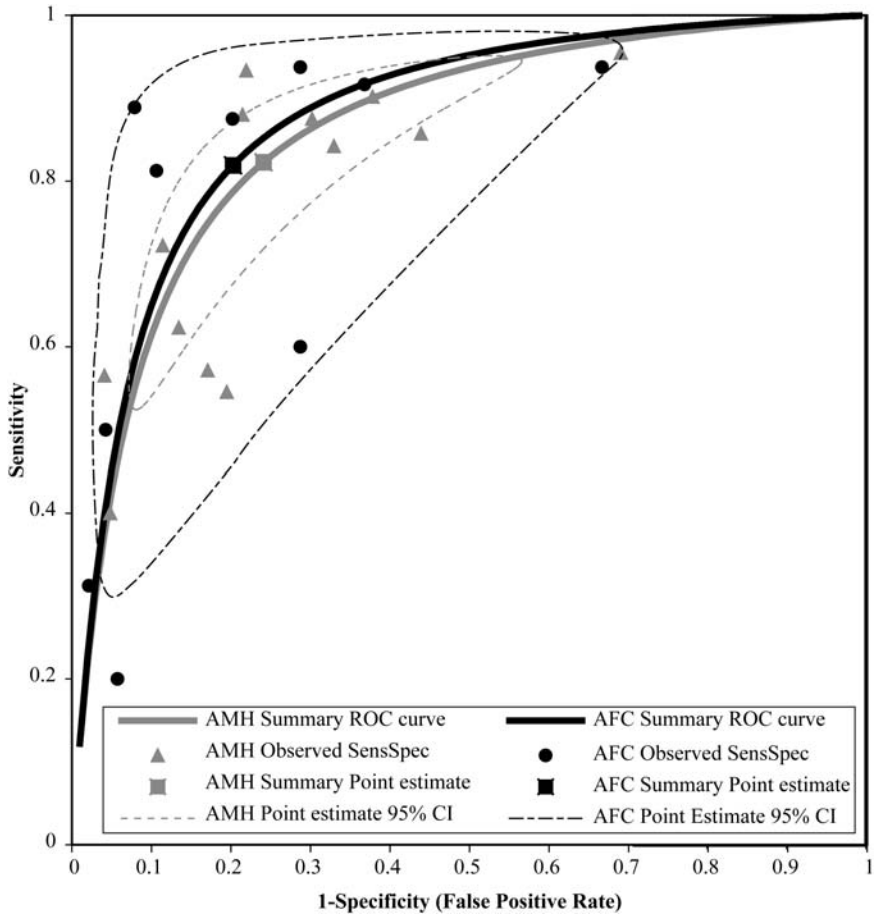

Figure 2 AMH and AFC in the prediction of an excessive response. *Note:* Regardless of the number of cut-offs mentioned per study, only one cut-off was taken into analysis. For the observed values of sensitivity–specificity points, all cut-offs are displayed.

deciding who will be a predicted excessive responder. But, AMH is a cycle independent test (Cook *et al*., 2000; Hehenkamp *et al*., 2006; La Marca *et al*., 2006; Tsepelidis *et al*., 2007), so any measurement in the period before starting the ART cycle will be at the disposition of the clinician, making the test an ideal tool. For the AFC, standardization needs to be dealt with by the physician (Broekmans *et al*., 2009), implying choices on ultrasound equipment, dedicated personnel and a systematic visualization and counting process. As the intra- and between cycle stability for the AFC may be comparable to that for AMH (van Disseldorp *et al*., 2010), the unlimited availability of this test makes it the preferable one for the short-term.

Currently, the identification of patients at risk for excessive response is based on a variety of factors, such as age, body weight and the presence of polycystic ovaries (Kwee *et al*., 2007; Lee *et al*., 2008; Riggs *et al*., 2008). However, the predictive value of these factors is quite poor. Whether their addition to tests like the AFC and AMH will improve the predictive capacity in identifying excessive responders remains to be established. In fact, patients with the polycystic ovary syndrome have clearly elevated AFC's and AMH levels (Broekmans *et al*., 2008; La Marca *et al*., 2009). As most studies have not excluded PCOS cases, these cases will add to the current analysis. In studies on PCOS cases only, AMH levels do indeed predict ovarian response to controlled hyperstimulation (Kaya *et al*., 2010). A limited number of studies exist on the use of multifactor prediction of the number of oocytes retrieved, using female age, basal FSH, ovarian ultrasound and smoking behaviour as predictors (Popovic-Todorovic *et al*., 2003; Howles *et al*., 2006). Validation of these prediction models in external populations has not been carried out so far. Individual patient data analyses of published literature may enable the assessment of the true value of such multivariable approach, combining patient characteristics and test results (Broeze *et al*., 2009).

Implications for clinical practice

Excessive response to ovarian stimulation induces the risk of the OHSS, especially in cases where exaggerated response is followed by a pregnancy. It may also cause increased patient discomfort and even reduced prospects for pregnancy. Up to 30% of IVF cycles are accompanied by complaints of mild or moderate OHSS and in 3–8% the severe form of OHSS may develop (Delvigne and Rozenberg, 2002). Once an excessive response has occurred, hCG administration could be withheld in an effort to eliminate this risk. Protective measures have also been reported for conditions in which oocyte retrieval has been allowed. Albumin administered at

Table III Clinical value of AMH and AFC in the prediction of an excessive response.

Prediction of an excessive response (Pre-test probability 20%)			
LR range	Occurrence of abnormal test result in LR range (%)		Post-test probability of excessive response (%)
	AMH	AFC	
<2	44	43	20–33
2–3	16	16	33–43
3–4	9	8	43–50
4–5	5	5	50–56
5–6	5	5	56–60
6–7	4	4	60–64
7–8	3	3	64–67
>8	14	16	>67

The occurrence of both AMH and AFC results within a specified likelihood ratio (LR) range and the concomitant post-test probabilities of an excessive response are shown, given a prevalence of an excessive response of 20%. For example, at a positive likelihood ratio (LR+) of at least 6, the post-test probability is 60% if the prior chance of having an excessive response is 20%. With the cut-off levels for the test corresponding to these LR+ levels the proportion of abnormal tests is 21% for the AMH and 23% for the AFC.

the time of oocyte retrieval, elective cryopreservation of all embryos to prevent the occurrence of pregnancy in the fresh cycle, GnRH agonist use for endogenous LH triggered ovulation in gonadotrophins/GnRH antagonist cycles and of the use of a single-dose recombinant LH to trigger ovulation have all been proposed (Aboulghar, 2009; Busso *et al.*, 2009; Kol and Dor, 2009; Kosmas *et al.*, 2009). These measures may limit collateral damage linked to excessive response, but they certainly do not offer absolute protection. The prevention of excessive ovarian response may be considered the corner stone of preventive management for OHSS, as such responses add heavily to the risk of developing the syndrome (Aboulghar and Mansour, 2003).

The reduction of pregnancy chances in excessive responders is most likely caused by detrimental effects of concomitant supraphysiological hormone levels on oocytes and embryo quality (Simon *et al.*, 1995; Ertzeid and Storeng, 2001;Pena *et al.*, 2002). Moreover, exaggerated and untimely estrogen and progesterone concentrations will affect the orderly proliferation and subsequent luteinization of the endometrium and thereby its receptivity (Bourgain and Devroey, 2003; Devroey *et al.*, 2004; Kodaman and Taylor, 2004). Moreover, an excessive ovarian response results in the yield of additional immature oocytes that are likely to be of insufficient quality to result in conception (Baart *et al.*, 2006; van der Gaast *et al.*, 2006).

Prior information on the expected ovarian response may allow the application of individualized stimulation protocols that will mitigate the number of follicles growing. The ideal test for excessive response prediction would identify all women with an excessive response and exclude all those women with a normal or poor response to standard dose stimulation. These women could be given individualized, milder treatment regimens, ensuring a yield of oocytes between 5 and 12 oocytes (Nargund *et al.*, 2007). In reality, tests like the AFC and AMH will never be absolutely accurate in their prediction. Assuming that a cut-off can be used at which 75% of excessive responders will be identified, a considerable number of excessive responders will be turned into normal responders by using for instance a lower than standard dose of FSH. At the same time, the abnormal test will be falsely positive in some 15% of cases, and a lower dose may turn these cases into poor responders. Whether this 'poor' response may alter the prospects for pregnancy may be disputed, as mild stimulation protocols have demonstrated that in normal profile cases, a mild response does not affect outcome (Out *et al.*, 2001; Heijnen *et al.*, 2007; Olivennes *et al.*, 2009; Verberg *et al.*, 2009).

Currently, only a few studies have addressed the use of reduced dosages of FSH based on prior prediction of the ovarian response level. In the study by Popovic-Todorovic *et al.* (2003), individualized FSH dosing appeared not to reduce the proportion of excessive responders, although the reduced dose group produced on average two oocytes less than the standard stimulated group. In one other study, Olivennes *et al.*, demonstrated that in predicted excessive responders, the use of FSH dosages lower than 150 IU produced mild ovarian responses without compromising pregnancy rates. A randomized comparison of standard versus individualized treatment based on the CONSORT prediction algorithm has recently been finalized and results are awaited (Olivennes *et al.*, 2009). Such studies should not only focus on the achievement of a more homogenous ovarian response. Also cost-economic effects regarding prevention of severe OHSS and a reduction of FSH consumption will aid in rationalising ovarian stimulation protocols for IVF.

Limitations

Although the process of systematic literature review and meta-analysis is a practical way to generate a more powerful estimate of true effect-size with less random error than individual studies, it does come with some limitations. First of all, the heterogeneity of studies must be addressed, as it may affect the justification for pooling the data into one analysis. In the case of the present meta-analysis, heterogeneity was caused by both different study quality characteristics and slight differences in study populations. Additionally, the definition of excessive response was not uniform across studies (Table I) and varied from the use of a threshold for number of oocytes aspirated to the development of OHSS. Another limitation was the allocation, by the authors, of different cut-off values for AMH and AFC. This is problematic as it interferes with the identification of a single threshold for AMH of AFC that could be predictive of an excessive response. The solution for this problem is the construction of a ROC curve, by which the effect of different cut-offs on the sensitivity/specificity combinations will become clear and overall accuracy becomes apparent.

Many of these methodological problems may be overcome by using individual patient database meta-analysis. From such data sets, population and patient characteristics, test results, stimulation data and outcome variables can be uniformed as much as possible before applying meta-analysis. Currently, initiatives in this field have been employed (Broeze *et al.*, 2009).

Lastly, there are some limitations that apply specifically to the method used to assess AMH levels and the AFC. The studies in this meta-analysis did not all use the same AMH assay. There is a noteworthy difference between the Beckmann–Coulter ELISA and the Diagnostic System Laboratories (DSL) ELISA leading to a wide dispersion of AMH

concentrations (Freour *et al*., 2007). This compatibility problem can only be overcome by the development of an internationally standardized AMH assay (Freour *et al*., 2007). Similar problems arise with the use of AFC results, where either follicle sizes of 2–5 or 2–10 mm are included into the counts. Although both methods of measurement may deliver the same level of accuracy for the test, it certainly will hamper the identification of a generally applicable cut-off.

Future research

The role of ovarian reserve tests in excessive response prediction combined with simple patient characteristics could be further analysed by using large individual patient data sets. The EXPORT (individual meta-analysis of patient data for Excessive Response Prediction with Ovarian Reserve Tests) initiative may offer the opportunity to start such effort. Moreover, stimulation protocols tailored on the basis of ovarian response prediction should be analysed as to their effects on pregnancy rates, costs for medication and patient satisfaction. Only large randomized comparisons of standard treatment strategies versus individualized treatment approaches will provide the correct answers, and will enforce previous undertakings in this area (Popovic-Todorovic *et al*., 2003).

Summary

The current systematic review and meta-analysis demonstrates that both the AFC and AMH are capable of identifying excessive responders to ovarian stimulation for IVF. Test optimization for clinical application may be more promising for AMH.

Conflict of interest: none declared.

Funding

F.J.M.B. is a member of the external advisory board for Ferring Pharmaceuticals, Hoofddorp, The Netherlands. He receives no monetary compensation. B.C.F. has received fees and grant support from the following companies (in alphabetic order); Andromed, Ardana, Ferring, Genovum, Merck Serono, Organon, Pantharei Bioscience, PregLem, Schering, Schering Plough, Serono and Wyeth.

References

Aboulghar M. Symposium: update on prediction and management of OHSS. Prevention of OHSS. *Reprod Biomed Online* 2009;**19**:33–42.

Aboulghar MA, Mansour RT. Ovarian hyperstimulation syndrome: classifications and critical analysis of preventive measures. *Hum Reprod Update* 2003;**9**:275–289.

Aflatoonian A, Oskouian H, Ahmadi S, Oskouian L. Prediction of high ovarian response to controlled ovarian hyperstimulation: anti-Mullerian hormone versus small antral follicle count (2–6 mm). *J Assist Reprod Genet* 2009;**26**:319–325.

Baart EB, Martini E, van dB I, Macklon NS, Galjaard RJ, Fauser BC, van Opstal D. Preimplantation genetic screening reveals a high incidence of aneuploidy and mosaicism in embryos from young women undergoing IVF. *Hum Reprod* 2006;**21**:223–233.

Baart EB, Martini E, Eijkemans MJ, Van OD, Beckers NG, Verhoeff A, Macklon NS, Fauser BC. Milder ovarian stimulation for in-vitro fertilization reduces aneuploidy in the human preimplantation embryo: a randomized controlled trial. *Hum Reprod* 2007;**22**:980–988.

Bersinger NA, Wunder D, Birkhauser MH, Guibourdenche J. Measurement of anti-mullerian hormone by Beckman Coulter ELISA and DSL ELISA in assisted reproduction: differences between serum and follicular fluid. *Clin Chim Acta* 2007;**384**:174–175.

Bourgain C, Devroey P. The endometrium in stimulated cycles for IVF. *Hum Reprod Update* 2003;**9**:515–522.

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.

Broekmans FJ, Visser JA, Laven JS, Broer SL, Themmen AP, Fauser BC. Anti-Mullerian hormone and ovarian dysfunction. *Trends Endocrinol Metab* 2008;**19**:340–347.

Broekmans FJ, de ZD, Howles CM, Gougeon A, Trew G, Olivennes F. The antral follicle count: practical recommendations for better standardization. *Fertil Steril* 2009. Epub ahead of print July 7.

Broer SL, Mol BW, Hendriks D, Broekmans FJ. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. *Fertil Steril* 2009;**91**:705–714.

Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PM, Coppus SF, Johnson NP, Khan KS, Ter RG, van der Veen F *et al*. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol* 2009;**9**:22.

Busso CE, Garcia-Velasco J, Gomez R, Alvarez C, Simon C, Pellicer A. Symposium: Update on prediction and management of OHSS. Prevention of OHSS—dopamine agonists. *Reprod Biomed Online* 2009;**19**:43–51.

Cook CL, Siow Y, Taylor S, Fallat ME. Serum mullerian-inhibiting substance levels during normal menstrual cycles. *Fertil Steril* 2000;**73**:859–861.

de Carvalho BR, Rosa e Silva AC, Rosa E Silva JC, dos Reis RM, Ferriani RA, Silva de Sa MF. Ovarian reserve evaluation: state of the art. *J Assist Reprod Genet* 2008;**25**:311–322.

Delvigne A, Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review. *Hum Reprod Update* 2002;**8**:559–577.

Devroey P, Bourgain C, Macklon NS, Fauser BC. Reproductive biology and IVF: ovarian stimulation and endometrial receptivity. *Trends Endocrinol Metab* 2004;**15**:84–90.

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**:2022–2026.

Eldar-Geva T, Ben Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, Gal M, Zylber-Haran E, Margalioth EJ. 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**:3178–3183.

Ertzeid G, Storeng R. The impact of ovarian stimulation on implantation and fetal development in mice. *Hum Reprod* 2001;**16**:221–225.

Fauser BC, Diedrich K, Devroey P. Predictors of ovarian response: progress towards individualized treatment in ovulation induction and ovarian stimulation. *Hum Reprod Update* 2008;**14**:1–14.

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**:162–164.

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;**8**:239–251.

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**:4057–4063.

Heijnen EM, Eijkemans MJ, de Klerk C, Polinder S, Beckers NG, Klinkert ER, Broekmans FJ, Passchier J, Te Velde ER, Macklon NS, Fauser BC. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. *Lancet* 2007;**369**:743–749.

Howles CM, Saunders H, Alam V, Engrand P. Predictive factors and a corresponding treatment algorithm for controlled ovarian stimulation in patients treated with recombinant human follicle stimulating hormone (follitropin alfa) during assisted reproduction technology (ART) procedures. An analysis of 1378 patients. *Curr Med Res Opin* 2006;**22**:907–918.

Kaya C, Pabuccu R, Satiroglu H. Serum antimullerian hormone concentrations on day 3 of the in vitro fertilization stimulation cycle are predictive of the fertilization, implantation, and pregnancy in polycystic ovary syndrome patients

undergoing assisted reproduction. *Fertil Steril* 2010. Epub ahead of print February 4.

Kodaman PH, Taylor HS. Hormonal regulation of implantation. *Obstet Gynecol Clin North Am* 2004;**31**:745–766. ix.

Kol S, Dor J. Symposium: Update on prediction and management of OHSS. Prevention of OHSS: GnRH agonist versus HCG to trigger ovulation. *Reprod Biomed Online* 2009;**19**:59–60.

Kosmas IP, Zikopoulos K, Georgiou I, Paraskevaidis E, Blockeel C, Tournaye H, Van Der EJ, Devroey P. Low-dose HCG may improve pregnancy rates and lower OHSS in antagonist cycles: a meta-analysis. *Reprod Biomed Online* 2009; **19**:619–630.

Kwee J, Schats R, McDonnell J, Themmen A, de Jong F, Lambalk C. Evaluation of anti-Müllerian hormone as a test for the prediction of ovarian reserve. *Fertil Steril* 2008;**90**:737–743.

La Marca A, Stabile G, Artenisio AC, Volpe A. Serum anti-Mullerian hormone throughout the human menstrual cycle. *Hum Reprod* 2006;**21**:3103–3107.

La Marca A, Giulini S, Tirelli A, Bertucci E, Marsella T, Xella S, Volpe A. Anti-Mullerian hormone measurement on any day of the menstrual cycle strongly predicts ovarian response in assisted reproductive technology. *Hum Reprod* 2007;**22**:766–771.

La Marca A, Broekmans FJ, Volpe A, Fauser BC, Macklon NS. Anti-Mullerian hormone (AMH): what do we still need to know? *Hum Reprod* 2009; **24**:2264–2275.

Lee TH, Liu CH, Huang CC, Wu YL, Shih YT, Ho HN, Yang YS, Lee MS. Serum anti-Mullerian hormone and estradiol levels as predictors of ovarian hyperstimulation syndrome in assisted reproduction technology cycles. *Hum Reprod* 2008;**23**:160–167.

Lekamge DN, Barry M, Kolo M, Lane M, Gilchrist RB, Tremellen KP. Anti-Mullerian hormone as a predictor of IVF outcome. *Reprod Biomed Online* 2007;**14**:602–610.

Nakhuda GS, Chu MC, Wang JG, Sauer MV, Lobo RA. Elevated serum mullerian-inhibiting substance may be a marker for ovarian hyperstimulation syndrome in normal women undergoing in vitro fertilization. *Fertil Steril* 2006; **85**:1541–1543.

Nakhuda GS, Sauer MV, Wang JG, Ferin M, Lobo RA. Mullerian inhibiting substance is an accurate marker of ovarian response in women of advanced reproductive age undergoing IVF. *Reprod Biomed Online* 2007;**14**:450–454.

Nardo LG, Gelbaya TA, Wilkinson H, Roberts SA, Yates A, Pemberton P, Laing I. Circulating basal anti-Mullerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. *Fertil Steril* 2009;**92**:1586–1593.

Nargund G, Fauser BC, Macklon NS, Ombelet W, Nygren K, Frydman R. The ISMAAR proposal on terminology for ovarian stimulation for IVF. *Hum Reprod* 2007;**22**:2801–2804.

Nelson SM, Yates RW, Fleming R. Serum anti-Mullerian hormone and FSH: prediction of live birth and extremes of response in stimulated cycles–implications for individualization of therapy. *Hum Reprod* 2007;**22**:2414–2421.

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**:1937–1942.

Olivennes F, Howles CM, Borini A, Germond M, Trew G, Wikland M, Zegers-Hochschild F, Saunders H, Alam V. Individualizing FSH dose for assisted reproduction using a novel algorithm: the CONSORT study. *Reprod Biomed Online* 2009;**18**:195–204.

Out HJ, David I, Ron-El R, Friedler S, Shalev E, Geslevich J, Dor J, Shulman A, Ben Rafael Z, Fisch B *et al.* A randomized, double-blind clinical trial using fixed daily doses of 100 or 200 IU of recombinant FSH in ICSI cycles. *Hum Reprod* 2001; **16**:1104–1109.

Pena JE, Chang PL, Chan LK, Zeitoun K, Thornton MH, Sauer MV. Supraphysiological estradiol levels do not affect oocyte and embryo quality in oocyte donation cycles. *Hum Reprod* 2002;**17**:83–87.

Popovic-Todorovic B, Loft A, Lindhard A, Bangsboll S, Andersson AM, Andersen AN. A prospective study of predictive factors of ovarian response in 'standard' IVF/ICSI patients treated with recombinant FSH. A suggestion for a recombinant FSH dosage normogram. *Hum Reprod* 2003;**18**:781–787.

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–990.

Riggs RM, Duran EH, Baker MW, Kimble TD, Hobeika E, Yin L, Matos-Bodden L, Leader B, Stadtmauer L. Assessment of ovarian reserve with anti-Mullerian hormone: a comparison of the predictive value of anti-Mullerian hormone, follicle-stimulating hormone, inhibin B, and age. *Am J Obstet Gynecol* 2008; **199**:202–208.

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**: 468–471.

Simon C, Cano F, Valbuena D, Remohi J, Pellicer A. Clinical evidence for a detrimental effect on uterine receptivity of high serum oestradiol concentrations in high and normal responder patients. *Hum Reprod* 1995; **10**:2432–2437.

Tremellen KP, Kolo M, Gilmore A, Lekamge DN. Anti-mullerian hormone as a marker of ovarian reserve. *Aust N Z J Obstet Gynaecol* 2005;**45**:20–24.

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**:1837–1840.

van Disseldorp J, Lambalk CB, Kwee J, Looman CW, Eijkemans MJ, Fauser BC, Broekmans FJ. Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. *Hum Reprod* 2010;**25**:221–227.

van der Gaast MH, Eijkemans MJ, van der Net JB, de Boer EJ, Burger CW, van Leeuwen FE, Fauser BC, Macklon NS. Optimum number of oocytes for a successful first IVF treatment cycle. *Reprod Biomed Online* 2006;**13**:476–480.

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

Verberg MF, Macklon NS, Nargund G, Frydman R, Devroey P, Broekmans FJ, Fauser BC. Mild ovarian stimulation for IVF. *Hum Reprod Update* 2009;**15**: 13–29.