Human Reproduction Update, Vol.19, No.1 pp. 26–36, 2013
Advanced Access publication on November 27, 2012 doi:10.1093/humupd/dms041

human reproduction update

Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach

Simone L. Broer[1,2,*‡], Jeroen van Disseldorp[1,2‡], Kimiko A. Broeze[1,2], Madeleine Dolleman[1,2], Brent C. Opmeer[1,2], Patrick Bossuyt[1,2], Marinus J.C. Eijkemans[1,2], Ben-Willem J. Mol[1,2], and Frank J.M. Broekmans[1,2] on behalf of the IMPORT study group[**]

[1]Department of Reproductive Medicine, University Medical Center Utrecht, Room F05.126, P.O. Box 85500, Utrecht 3508 GA, The Netherlands [2]Department of Obstetrics and Gynaecology, Amsterdam Medical Center, Amsterdam, the Netherlands

*Correspondence address. Tel: +31-88-755-1044; Fax: +31-88-755-5433; E-mail: S.L.Broer-2@umcutrecht.nl, www.ipd-meta-analysis.com/ipd
**A full list of authors is provided in the Appendix section.

Submitted on November 5, 2011; resubmitted on July 10, 2012; accepted on July 30, 2012

TABLE OF CONTENTS

- Introduction
- Methods
  - Data acquisition
  - Statistical analysis
- Results
  - Data acquisition
  - Prediction of a poor response or ongoing pregnancy from patient characteristics
  - Prediction of a poor response or ongoing pregnancy from ovarian reserve tests
  - Multivariable prediction models for poor response and ongoing pregnancy
  - Accounting for FSH dosage and study quality
- Discussion

BACKGROUND: Although ovarian reserve tests (ORTs) are frequently used prior to IVF treatment for outcome prediction, their added predictive value is unclear. We assessed the added value of ORTs to patient characteristics in the prediction of IVF outcome.

METHODS: An individual patient data (IPD) meta-analysis from published studies was performed. Studies on FSH, anti-Müllerian hormone (AMH) or antral follicle count (AFC) in women undergoing IVF were identified and authors were contacted. Using random intercept logistic regression models, we estimated the added predictive value of ORTs for poor response and ongoing pregnancy after IVF, relative to patient characteristics.

RESULTS: We were able to collect 28 study databases, comprising 5705 women undergoing IVF. The area under the receiver-operating characteristic curve (AUC) for female age in predicting poor response was 0.61. AFC and AMH each significantly improved the model fit (*P*-value <0.001). Moreover, almost a similar accuracy was reached using AMH or AFC alone (AUC 0.78 and 0.76, respectively). Combining

‡ Both authors contributed equally.

the two tests, however, did not improve prediction (AUC 0.80, $P = 0.19$) of poor response. In predicting ongoing pregnancy after IVF, age was the best single predictor (AUC 0.57), and none of the ORTs added any value.

CONCLUSIONS: This IPD meta-analysis demonstrates that AFC and AMH clearly add to age in predicting poor response. As single tests, AFC and AMH both fully cover the prediction of poor ovarian response. In contrast, none of the ORTs add any information to the limited capacity of female age to predict ongoing pregnancy after IVF. The clinical usefulness of ORTs prior to IVF will be limited to the prediction of ovarian response.

Key words: ovarian reserve tests / AMH / AFC / individual patient data meta-analysis / IVF outcome prediction

Introduction

The incorporation of ovarian reserve tests (ORTs) in IVF management started after initial publications indicated a potential role for basal FSH in predicting pregnancy outcome after IVF and in counseling patients (Muasher *et al.*, 1988; Scott *et al.*, 1989). Since these first publications, a large body of additional work on basal FSH and several other tests has been published, often with inconsistent findings on the magnitude and direction of the predictive effect. It became evident that the clinical value of previously published prediction models was highly dependent on the consequences related to the prediction (i.e. counseling versus refraining from treatment). Moreover, female age, which is strongly related to IVF outcome, was frequently omitted as a prime contributor in the prediction models (Broekmans *et al.*, 2006; Verhagen *et al.*, 2008).

Overall, individual studies have shown considerable variation in the predictive capacity of ORTs. The conventional way to summarize the available evidence would be to perform a systematic review and meta-analysis of the sensitivity and specificity of ORT, as reported in published studies (Leeflang *et al.*, 2008). However, a major problem in interpreting these studies is the striking heterogeneity in individual patient populations, stimulation protocols, hormone assays ultrasound techniques and other features. Conventional meta-analysis of the accuracy of tests cannot easily account for this heterogeneity, nor does it respect the continuous nature of ORT data, or the statistical dependence between related tests and variables, i.e. ORT results are related to female age, and both are predictive of IVF outcome (Broeze *et al.*, 2009).

To arrive at summary estimates of the added value of ORTs in women undergoing IVF, we undertook a meta-analysis with original individual patient data (IPD). By collecting test results, age and other patient characteristics, and IVF outcome in each individual patient, we would be able to respect the continuous nature of ORT data and would be able to study the added value of ORT to basic patient characteristics in predicting IVF outcome. Our aim was to answer the question of whether the most widely used ORTs, follicle stimulating hormone (FSH), antral follicle count (AFC) and anti-Müllerian hormone (AMH) added significantly and substantially to baseline female characteristics, such as age, in predicting the outcome of IVF treatment.

Methods

Data acquisition

We searched the literature for studies on the value of FSH, AFC and AMH in predicting IVF outcome. We built on searches performed in previous, conventional systematic reviews on the subject (Broekmans *et al.*, 2006; Broer *et al.*, 2009). A systematic search was performed in Medline to identify additional eligible papers, published until December 2009 (Fig. 1). Eligible for the current review were studies presenting data on at least one ORT and at least one patient characteristic and IVF outcome, in terms of ovarian response to stimulation, clinical or ongoing pregnancy or both.

Keywords used were synonyms for *in vitro* fertilization (IVF, controlled ovarian stimulation, *in vitro* fertilization) and synonyms for the respective ORT (FSH, Follicle Stimulating Hormone, AFC, Antral Follicle Count or number, AMH, Anti-Müllerian Hormone, Müllerian inhibiting substance). All titles and abstracts were evaluated for eligibility by two authors (S.B., J.D.) and if necessary the opinion of a third author was decisive (F.B.).

All authors of identified potentially eligible primary studies were informed about this IPD meta-analysis project and invited to share their data in a collaborative project. If authors were inclined to participate, they were provided with a data request form, informing them of the format of the data requested.

After data acquisition, all data were carefully examined and when possible converted into a single format. Any issues or inconsistencies were checked with the original author. For a more detailed description of our IPD meta-analysis methodology, the reader is referred to previous papers of our group (Broeze *et al.*, 2009, 2010).

A comparison was made between the studies that could and those that could not be included. Sensitivity and specificity pairs at a certain threshold level for the prediction of a poor response or ongoing pregnancy were calculated for included and not included studies. A Spearman correlation was then calculated for sensitivity and specificity to specify that the differences in sensitivity and specificity levels between included and not included studies were only the result of the different threshold levels used.

We evaluated the quality of the included studies using the QUADAS checklist, supplemented by a number of items to evaluate the risk of bias in prognostic studies.

Statistical analysis

All analyses were performed for predicting both poor response as well as ongoing pregnancy after IVF as the outcome of interest. Poor response was defined as the yield of four or fewer oocytes at follicle aspiration or a cancelled cycle due to poor ovarian response ($<3–4$ dominant follicles >12 mm diameter), since this is a commonly used definition for poor response (Broer *et al.*, 2009). Ongoing pregnancy was defined as a visible gestational sac on ultrasound with heartbeat at a gestational age of at least 9 weeks.

Duration of subfertility was defined as the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt. Patients were stimulated according to the local protocol. In almost all studies, a starting dosage of at least 150 IU of FSH was used, which is the optimal daily dosage in expected normal responders. With this dose, it can be assumed that all patients receive maximum stimulation,

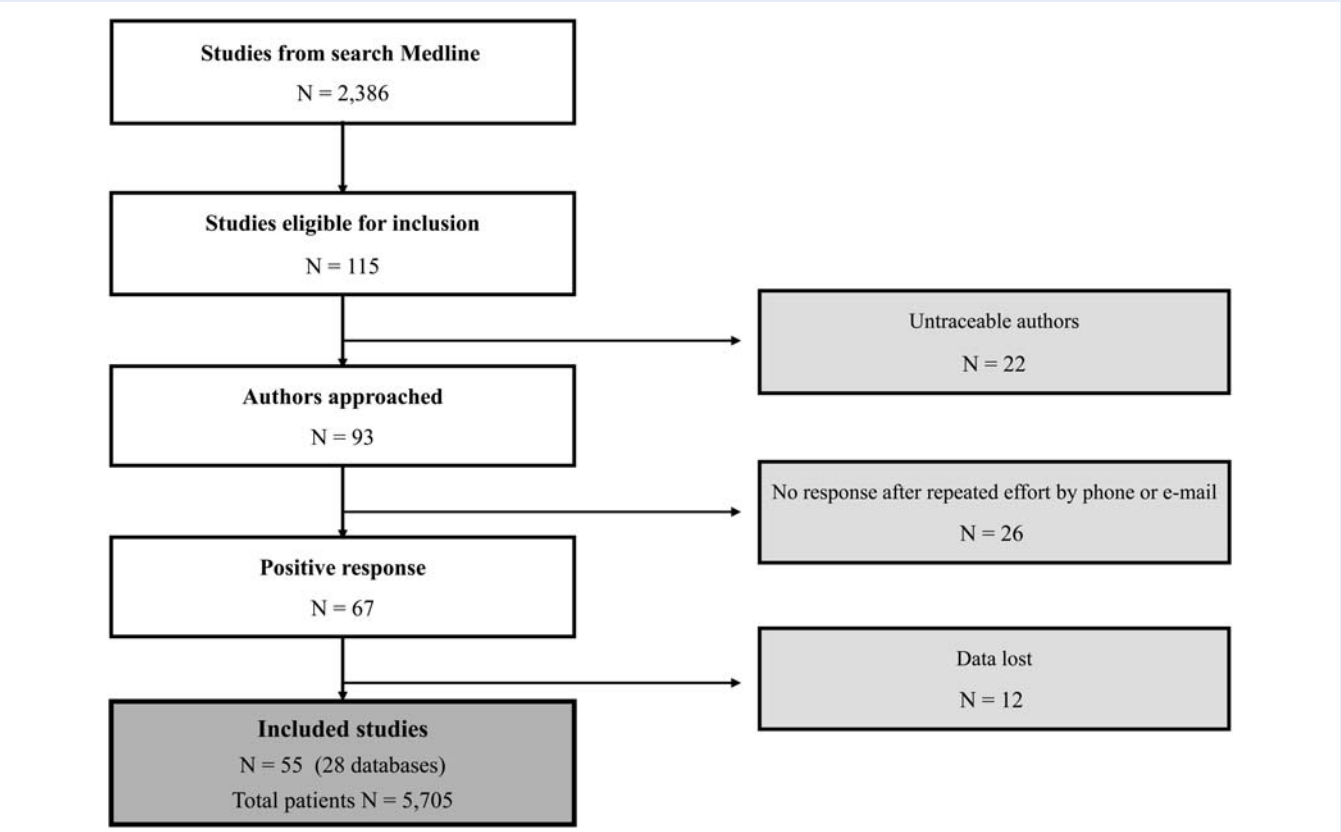

Figure 1 Flowchart of included studies.

creating growth of all follicles sensitive to FSH in that time frame (Sterrenburg *et al.*, 2011). We evaluated whether the ORT and patient characteristics, female age, BMI and duration of subfertility, were missing in the individual study databases. Whenever a particular variable was missing in an individual database, we made no attempt at imputation.

Random intercept logistic regression prediction models were then created with the 'Lme4' library in R version 2.9.0. (http://www.r-project.org/), using the Laplace approximation to the likelihood. These models were created to quantitatively estimate the added value of the ORTs on the patient characteristics in predicting poor response or ongoing pregnancy.

The random intercept model takes heterogeneity into account by assuming that included studies are a random sample of a potential universe of studies and that between-study variation in the predictive effect in this universe can be described by a normal distribution on the log odds scale. The model provides an estimate of the summary predictive effect as well as of the standard deviation of this distribution.

Three different sets of models for the prediction of poor response or ongoing pregnancy were used. The first model set included the patient characteristics female age, BMI and duration of subfertility. In the second set of models, the predictive capacities of the individual ORTs, FSH, AFC or AMH, in combination with predictive patient characteristics, were estimated. In the third set of models, the added value of combinations of ORTs to the patient characteristics was evaluated.

To account for between-study differences and their potential effect on our conclusions, we repeated the analyses using starting FSH dosage as a covariate. In similar analyses, we included study design features, as identified by the QUADAS checklist, as covariates in our models, to evaluate whether accounting for study design differences attenuated the observed associations between ORT, patient characteristics and the respective outcomes.

We constructed receiver-operating characteristic (ROC) curves to express the predictive accuracy of each of the combinations of predictive variables: the ability of the model to distinguish poor responders from the rest, and the ability to distinguish successful IVF couples from the rest. With each of the random intercept logistic regression models, we calculated the probability of the outcome of interest (poor response or ongoing pregnancy). By moving a positivity threshold from 0 to 1, we could then calculate sensitivity–specificity pairs for each model. Based on these, we plotted stratified ROC curves, with the ROC regression model as proposed by Janes *et al.* (2009) and Pepe *et al.* (2009). This model assumes that studies share a common ROC but allows the positivity threshold corresponding to each sensitivity–specificity pair to vary between studies. With this model the improvement in predictive accuracy of adding an ORT to other variables can be studied, while correcting for the heterogeneity between studies. This way we could compare the ROC and area under the curves (AUCs) of the models described above and evaluate them for statistically significant differences.

Because not all studies in this meta-analysis had reported data for all three ORTs, we constructed the prediction models using those databases from the total dataset that included the three ORTs (FSH, AFC and AMH) and age, to allow for direct comparisons, minimizing bias from indirect comparisons. The results of our analyses in the subgroup of three-test studies are shown in the main text, the result of the total study group are shown in the Supplementary Files.

Data were analyzed using SPSS 17.0 (SPSS Inc., Chicago, IL, USA), SAS 9.1 (SAS Institute Inc., Cary, NC, USA) and R version 2.9.0. (http://www.r-project.org/).

Results

Data acquisition

We identified 115 eligible study reports, from which we obtained contact information from 93 authors. Of these 93 authors, 67 replied to our (repeated) email or phone contact. Ultimately, we received 28 study databases that had been used for the publication of 55 manuscripts, provided by 24 collaborating authors (Tomas *et al.*, 1997; Bancsi *et al.*, 2000; Ng *et al.*, 2000, 2005; Smeenk *et al.*, 2000; van Rooij *et al.*, 2002; Kwee *et al.*, 2003; Popovic-Todorovic *et al.*, 2003a, b; Yong *et al.*, 2003; Erdem *et al.*, 2004; Muttukrishna *et al.*, 2004, 2005; Vladimirov *et al.*, 2004, 2005; Ashrafi *et al.*, 2005; van Swieten *et al.*, 2005; Eldar-Geva *et al.*, 2005; Klinkert *et al.*, 2005a, b; Caroppo *et al.*, 2006; Jayaprakasan *et al.*, 2007; La Marca *et al.*, 2007; Luna *et al.*, 2007; McIlveen *et al.*, 2007; Merce *et al.*, 2007; Nelson *et al.*, 2007; Smeenk *et al.*, 2007; Liu and Greenblatt, 2008). These 28 databases contained data on 5705 subfertile women (Fig. 1). Data from 4170 women were suitable for poor response analysis, of these 893 (21%) had a poor response. Data from 5367 women could be used for the analysis of ongoing pregnancy prediction, of these 1231 women (23%) obtained an ongoing pregnancy.

Baseline characteristics of the 5705 women in the study group are summarized in Table I. The baseline characteristics of the original studies show some variation between the original studies, as do the poor response and pregnancy incidences, and the ORT averages (Supplementary data, Fig. S1). Study quality characteristics as scored by the QUADAS checklist and supplemental questions are shown in Fig. 2.

With the original data, we were able to replicate the primary findings of the original study in 10 databases. In 11 databases, the study database we received contained a number of patients that differed from the publication, whereas in 7 other databases there were slight inconsistencies in the baseline data previously published. The level of consistency between the individual data and the data reported in the published manuscript was considered sufficient for all included studies. No significant differences were found between the Spearman correlations of the included and not included studies for each ORT and outcome measure, indicating that the included and not included studies were comparable.

For 3235 women, both outcome measures were available and ongoing pregnancy rates for poor and normal responders could be compared. In normal responders, 30.2% of the women achieved an ongoing pregnancy compared with 11.7% of poor responders. Differences in ongoing pregnancy rates were also compared in age categories, demonstrating that poor responders have lower pregnancy rates than normal responders across all age groups, although this effect was gradually smaller in the higher age groups; over the age of 40 years comparable pregnancy prospects could be observed (Supplementary data, Table SI).

Prediction of a poor response or ongoing pregnancy from patient characteristics

For model building, we could use the data from 617 women for poor response analysis and from 420 women for ongoing pregnancy analysis. Of all patient characteristics, age was the strongest single predictor of poor response [odds ratio (OR) 1.12: 95% confidence interval (CI) 1.08–1.17; Supplementary data, Table SII]. BMI and duration of subfertility were not significantly predictive of poor response. In pregnancy prediction, age was the strongest single predictor of pregnancy, compared with other patient characteristics (OR 0.94: 95% CI 0.89–0.99) (Supplementary data, Table SII). Duration of subfertility was found not to be significantly associated with ongoing pregnancy, but BMI was. In a multivariable model, only BMI added any predictive value to age (Supplementary data, Table SII). Since age was the single constant and strongest predictor of poor response and ongoing pregnancy, we evaluated the added predictive effect of the ORTs FSH, AFC and/or AMH relative to the predictive value of age only in all further multivariable analyses.

Table I Baseline characteristics of the 5705 women in the study group.

	Mean (5th–95th percentile)
Patient characteristics	
Female age (years)	34.3 (26.7–41.9)
FSH (IU/l)	7.8 (3.8–14.0)
AFC (number)	11.6 (3.0–25.0)
AMH (ng/ml)	2.1 (0.1–6.0)
BMI (kg/m^2)	23.2 (18.5–30.1)
Duration of subfertility (years)	4.01 (1.0–9.1)
Prevalences	
Poor response	21.4%
Ongoing pregnancy	22.9%

Poor response: ≤4 oocytes retrieved. Ongoing pregnancy: positive heartbeat at AD >9 weeks. Duration of subfertility: the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt.

Prediction of a poor response or ongoing pregnancy from ORTs

We compared the ORTs using the random intercept logistic regression model in predicting poor response (Table II). The ROC regression analysis demonstrated a high accuracy for AMH (AUC 0.78: 95% CI 0.72–0.84) and for AFC (AUC 0.76: 95% CI 0.70–0.82) in predicting poor response, but only a moderate accuracy for FSH (AUC 0.68: 95% CI 0.61–0.74; Table III). In predicting pregnancy after IVF, all three ORT had only a very small or no predictive effect (Table II). The AUC were 0.53, 0.50 and 0.55 for FSH, AFC and AMH, respectively (Table III).

Multivariable prediction models for poor response and ongoing pregnancy

The multivariable analyses for poor response prediction showed that a model with age, AFC and AMH had a significantly higher predictive accuracy than a model based on age alone (AUC 0.80 versus 0.61; $P \leq 0.001$). Adding FSH to this model did not significantly improve predictive accuracy ($P = 0.45$; Table III). The predictive value of the multivariable model, including age and the two ORTs, AMH and AFC, was not significantly better than that of a single ORT, when used in isolation [$P = 0.17$ for AMH (AUC 0.78); $P = 0.99$ for AFC (AUC 0.76)]. AMH, as a single predictor, has a accuracy comparable with all multivariable models with AMH and

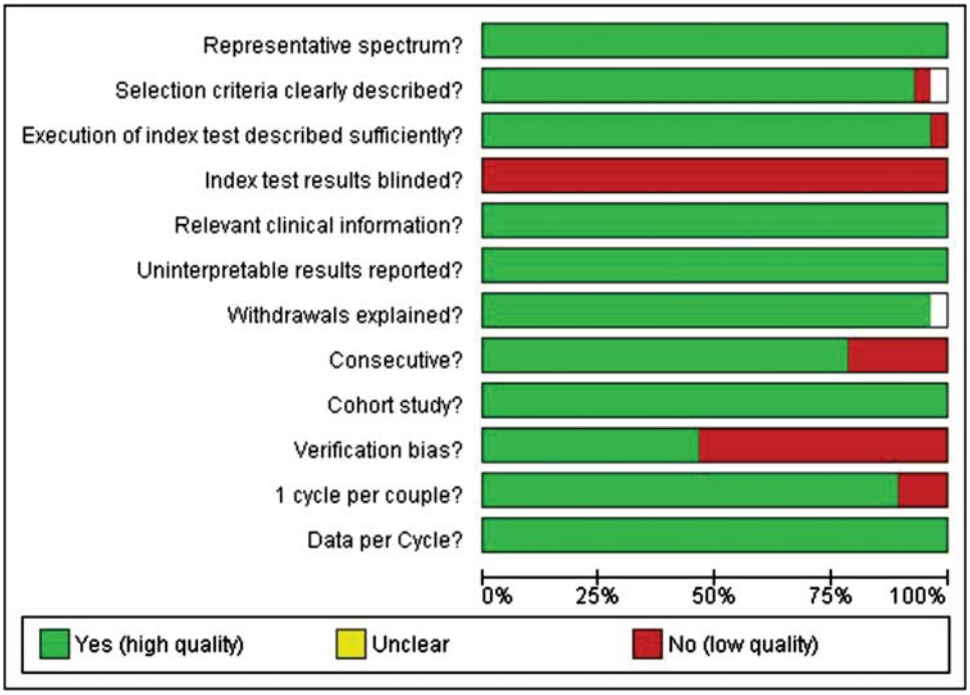

Figure 2 Study characteristics of the included studies. Characteristics of all included studies evaluated with the QUADAS checklist. Note that QUADAS was set up for diagnostic studies and these are all prognostic studies. Therefore all questions regarding the reference test could not be answered. Some questions specific for ovarian reserve testing and fertility studies were added. All studies were cohort studies, with the majority prospectively set up. All studies analyzed the results per cycle, some studies analyzed more cycles per couple, in which case only the first cycle was analyzed.

age or with any of the other two ORT. The ROC curves for the multivariable models are shown in Fig. 3A.

Age was the strongest single predictor of pregnancy after IVF, with moderate accuracy (AUC 0.57). Multivariable analysis for prediction of ongoing pregnancy indicated that no single or combined ORT significantly added predictive power to age (Table III). The AUC for the combination of age, AMH and AFC was 0.59. ROC curves for the multivariable analyses are shown in Fig. 3B.

Accounting for FSH dosage and study quality

In the prediction of a poor response, FSH dosage had a significant predictive effect (OR 1.009, $P < 0.001$). A higher FSH dosage was associated with higher chances of a poor response. When FSH dosage was added to the multivariable models of age and ORTs, the associations of age and ORTs with poor response were very similar to those of the models without accounting for FSH dosage. In the prediction of an ongoing pregnancy, FSH dosage did not have a significant effect (OR 0.997, P 0.140). Here also, inclusion in the multivariable models did not change the associations for age and ORTs. When included in the multivariable models, none of the evaluated study quality characteristics affected the predictive capacity of age and ORTs in predicting poor response or ongoing pregnancy.

Discussion

The results of this IPD meta-analysis, based on 28 studies previously reported, demonstrate that both AFC and AMH clearly add value to female age in the prediction of poor ovarian response in IVF. Comparably good predictions can be made with either AMH or AFC alone, without using female age. For the prediction of ongoing pregnancy after IVF, ORTs do not add to the limited predictive capacity of female age.

In the long-lasting debate on the true value of ovarian reserve testing prior to IVF, a systematic review of the literature with meta-analysis can be of help as an objective and systematic approach in summarizing the available evidence. A major strength of the collaborative effort reported here is its ability to analyze the independent added value of several relevant predictors in a large body of data. With the generous help of a large group of contributors, we have been able to collect data on a number of patients, which far surpasses that of the largest study performed so far, although it does not cover the entire evidence base. Thereby, we have achieved consistency in variable coding and a form of statistical analysis that accommodates the remaining heterogeneity between studies.

Some potential limitations of our study have to be acknowledged. For our analyses, the databases of only 55 of the eligible 115 manuscripts could be obtained. We were unable to reach a number of authors, primarily because of inaccurate contact information, or because authors did not reply on the e-mail addresses provided in the study reports. Furthermore, older data were often lost, or kept in a format that could no longer be read or could not be converted. The Spearman correlations of the included and not included studies were calculated and compared in order to study whether these groups of studies were comparable. For none of the ORTs with

Table II Univariable and multivariable models of age and ORT in the prediction of poor response and ongoing pregnancy.

	Poor response prediction				Ongoing pregnancy prediction			
	OR	95% CI	*P*-value	Variance RI	OR	95% CI	*P*-value	Variance RI
Univariable models								
Age (per year)	1.12	1.08–1.17	<0.001	0.412	0.94	0.89–0.99	0.011	0.441
FSH (per IU/l)	1.27	1.19–1.35	<0.001	0.559	0.98	0.92–1.04	0.477	0.537
AFC (per N)	0.77	0.73–0.82	<0.001	0.235	1.00	0.97–1.03	0.951	0.554
AMH (per ng/ml)	0.50	0.41–0.60	<0.001	0.440	1.09	0.96–1.24	0.197	0.462
Multivariable models								
Age and FSH				0.320				0.430
Age (per year)	1.12	1.07–1.17	<0.001		0.94	0.89–0.99	0.013	
FSH (per IU/l)	1.26	1.18–1.34	<0.001		0.99	0.93–1.05	0.632	
Age and AFC				0.192				0.476
Age (per year)	1.07	1.02–1.11	0.007		0.93	0.89–0.98	0.020	
AFC (per N)	0.78	0.74–0.83	<0.001		0.99	0.96–1.02	0.625	
Age and AMH				0.321				0.393
Age (per year)	1.08	1.03–1.13	0.001		0.94	0.89–0.99	0.017	
AMH (per ng/ml)	0.54	0.44–0.66	<0.001		1.06	0.93–1.21	0.373	

Results of random intercept logistic regression model in the prediction of poor response or ongoing pregnancy. For the prediction of a poor response, the multivariable analyses showed that all three ORT add predictive information to female age alone.
Female age is the strongest predictor of ongoing pregnancy. All three ORT show a very small or absent predictive effect in the prediction of an ongoing pregnancy. Multivariable analyses show that all three ORT do not add predictive information to female age alone in the prediction of an ongoing pregnancy. *P*-values reflect whether the variable plays a significant role in the model.
The column 'Variance RI' denotes the estimated variance of the random intercept in the random intercept logistic model. Its square root is the estimated standard deviation (SD), and may be interpreted on the logistic scale. A 1 SD difference in the population of studies corresponds to an increase in the Odds on the outcome (poor response and ongoing pregnancy, respectively) of exp (SD).

either outcome measure, a significant difference in the Spearman correlations was found. We therefore believe that the included and not included studies are comparable and that the current number of participants and level of detail allowed us to analyze a representative selection of the collected data.

The findings from our analyses confirm those of previous systematic reviews and meta-analysis of both single ORTs and multivariable prediction models for poor response to ovarian stimulation (Broekmans *et al.*, 2006; Verhagen *et al.*, 2008; Broer *et al.*, 2009). Both AMH and AFC strongly represent the size of the cohort of FSH sensitive follicles continuously present in the ovaries, often referred to as the quantitative ovarian reserve. Response to ovarian hyperstimulation has been shown to be directly linked to this cohort size (Kwee *et al.*, 2003). The role of AMH in marking the ovarian ageing process has been demonstrated in several studies showing that AMH decreases gradually with age and may be also predictive of the timing of menopause (de Vet *et al.*, 2002; van Rooij *et al.*, 2004, 2005; Sowers *et al.*, 2008; van Disseldorp *et al.*, 2008; Broer *et al.*, 2011). From these data, the capacity of AMH as a marker for the quantitative ovarian reserve has become established.

For ongoing pregnancy prediction, age is the single most important predictor, although the accuracy in pregnancy prediction remains far from optimal. In contrast to their performance in predicting poor response, ORTs perform poorly in predicting pregnancy, as shown in this analysis. The large body of data in the present analyses finally clarifies the lack of added value of currently known ORTs to knowing female age. Since ovarian response to controlled hyperstimulation reflects quantitative ovarian reserve and the occurrence of an ongoing pregnancy after IVF is mainly related to qualitative ovarian reserve, it can be emphasized that ORTs reflect the quantitative aspect of the ovarian reserve status only. Qualitative ovarian reserve appears much harder to evaluate. In addition, ovarian reserve may not be the only factor affecting pregnancy chances in IVF/ICSI. Several factors, such as embryo quality, transfer technique or endometrial receptivity may be important (Boomsma and Macklon, 2006). It is likely that only by studying several consecutive treatment cycles, a true representation of a woman's remaining reproductive capacity, based on her ovarian reserve status, may be obtained. Over the past decades, only one study evaluated the predictive role for ORTs in a series of subsequent IVF cycles, demonstrating that female age was the only factor predicting ongoing pregnancy after three treatment cycles, with no apparent role for ORTs (Hendriks *et al.*, 2005).

The performance of assisted reproduction technology (ART) in infertile couples is far from optimal. Out of every 100 couples initiating IVF, only 50–60 will achieve their goal, even after having undergone several treatment cycles (Lintsen *et al.*, 2007; Malizia *et al.*, 2009). This high failure rate could be attributed to several factors, of which drop out rates and reduced ovarian reserve are the most popular ones. The urge to improve ART performance puts a high focus on identifying adequate ORTs. The limited accuracy of current tests has led to the situation that unfavorable test outcomes only lead to counseling and treatment adaptations that lack a solid scientific basis, instead of a refusal to offer ART treatment in the first place. Recent

Table III AUCs of prediction models of age and ORTs for the prediction of a poor response and ongoing pregnancy.

	Three-test study group				Total study group			
	AUC	95% CI	*P*-value	*n*	AUC	95% CI	*P*-value	*n*
Poor response prediction								
Univariable models								
Age	0.61	0.54–0.68	NA	617	0.60	0.57–0.64	NA	4034
FSH	0.68	0.61–0.74	0.051	617	0.66	0.62–0.69	0.004	3652
AFC	0.76	0.70–0.82	<0.001	617	0.73	0.69–0.77	<0.001	2118
AMH	0.78	0.72–0.84	<0.001	617	0.81	0.77–0.84	<0.001	1274
Multivariable models								
Age and FSH	0.71	0.65–0.78	<0.001	617	0.69	0.66–0.72	<0.001	3652
Age and AFC	0.79	0.73–0.85	<0.001	617	0.76	0.72–0.80	<0.001	2118
Age and AMH	0.77	0.70–0.83	<0.001	617	0.80	0.76–0.84	<0.001	1274
Age and AMH and AFC	0.80	0.74–0.86	<0.001	617	0.80	0.74–0.86	<0.001	618
Age and AMH and AFC and FSH	0.81	0.75–0.86	<0.001	617	0.81	0.75–0.86	<0.001	617
Ongoing pregnancy prediction								
Univariable models								
Age	0.57	0.47–0.66	NA	420	0.56	0.54–0.59	NA	5207
FSH	0.53	0.43–0.62	0.348	420	0.54	0.51–0.58	0.084	3521
AFC	0.50	0.40–0.59	0.100	420	0.52	0.48–0.57	0.612	1977
AMH	0.55	0.45–0.64	0.630	420	0.58	0.51–0.64	0.495	1008
Multivariable models								
Age and FSH	0.58	0.48–0.67	0.195	420	0.60	0.57–0.64	0.116	3521
Age and AFC	0.58	0.48–0.67	0.247	420	0.57	0.52–0.61	0.709	1977
Age and AMH	0.57	0.48–0.67	0.753	420	0.59	0.53–0.65	0.415	1008
Age and AMH and AFC	0.59	0.49–0.68	0.371	420	0.59	0.49–0.68	0.341	421
Age and AMH and AFC and FSH	0.58	0.49–0.68	0.414	420	0.58	0.49–0.68	0.414	420

AUC, area under the curve; ORT, ovarian reserve test; AMH, anti-Müllerian hormone; AFC, antral follicle count; FSH, follicle stimulating hormone.
Poor response prediction. In the univariable analysis, it is shown that both AMH and AFC have a high accuracy, while FSH only has a moderate accuracy. In the multivariable models, the added value to the AUC of an ORT on female age is shown; the *P*-value indicates whether this added value is significant in comparison to age alone. All ORT show a significant rise in the AUC. Moreover, the added value of adding several ORTs to female age is shown. The model including age, AFC and AMH reached the maximum predictive power. This level of accuracy, however, is also obtained when using a two factor model in the total study group
Ongoing pregnancy. In the univariable analysis, it is shown that age is the strongest predictor compared with the single ORTs. The multivariable analysis shows that no single or combined ORT adds substantial predictive power to age alone. This is shown in the three tests study group, as well as in the total study group.

studies have suggested a role for the use of patient characteristics, especially female age, combined with AMH for identification of various prognosis categories, with the poorest group having a per-cycle chance of pregnancy of 5% with a rather wide precision interval (1–16%; La Marca *et al.*, 2011; Nelson and Lawlor, 2011). The question of how these predictions could alter patient management or aid in upgrading ART performance has remained unanswered. This may also be explained by the fact that very poor prognosis categories are difficult to identify with sufficient precision.

Recent publications have suggested the calculation of age-specific decline curves in order to maximize ORT accuracy (Barad *et al.*, 2007; Henne *et al.*, 2008; Scott *et al.*, 2008). One study calculated age-specific FSH levels and live birth probabilities and demonstrated that variation in the chances of live birth is primarily determined by age and only to a lesser degree by basal FSH (Henne *et al.*, 2008). The analysis also demonstrated that FSH decline curves for five age groups yielded different cut-off values in the prediction of delivery rates (Scott *et al.*, 2008). Since there was a very low rate of abnormal tests especially in the young age categories, the authors' conclusion that age-specific basal FSH testing could serve as a reliable prognostic tool may be too optimistic.

It is to be expected that similar issues in the evaluation of tests and markers can be resolved with the meta-analysis of IPD. More and more funding agencies are inviting investigators to have a data sharing policy, and to allow others to benefit from the resources invested in the research. Inspired by the major successes achieved by the multicenter genetic consortia, those interested in clinical research could develop similar initiatives for patient centered research. We strongly believe that joining efforts in multicenter collaborations, possible even fine-tuning and coordinating study protocols through prospective meta-analysis, is an inevitable next step for clinical science in the 21st century, not just for randomized trials of interventions, but also in the evaluation of medical tests and biomarkers.

The clinical use of markers like AMH, basal FSH and the AFC is mostly based on cut-off levels. From the individual patient dataset,

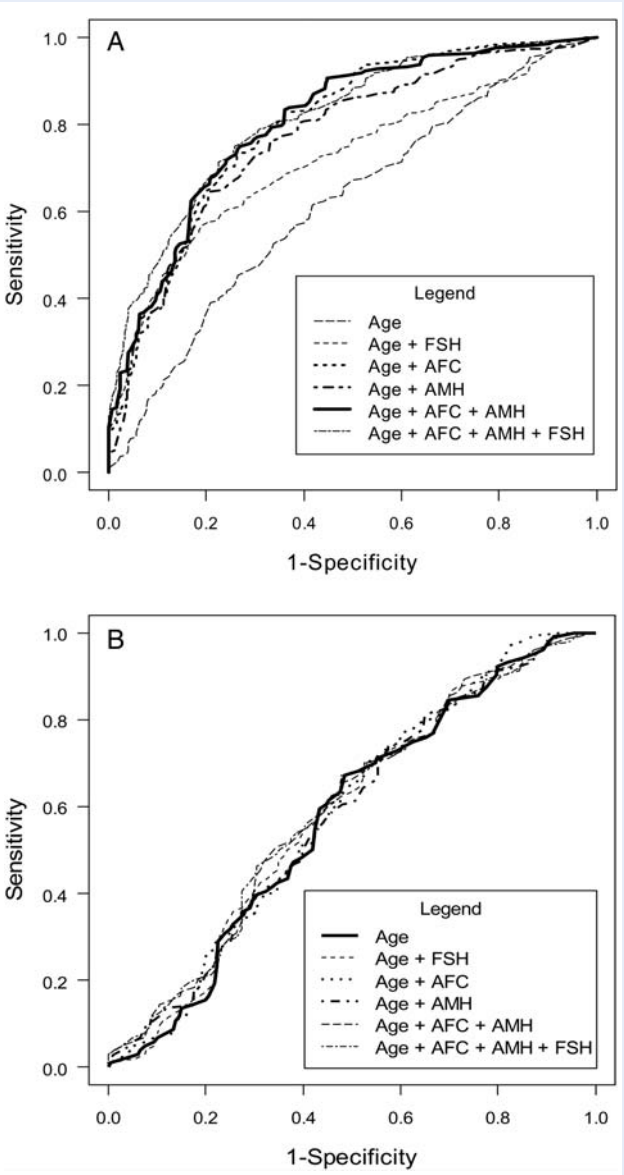

Figure 3 ROC curves of age and ORT in the prediction of poor response and ongoing pregnancy. (**A**) Poor response prediction based on age and ORT. The ROC curves of age or age combined with a single or more ORT are depicted. The ROC curves for 'Age + AMH', 'Age + AMH + AFC' and 'Age + AMH + AFC + FSH' run toward the upper left corner, indicating a good capacity to discriminate between normal and poor responders at certain cut-off levels. (**B**) Ongoing pregnancy prediction based on age and ORT. The ROC curves age or age combined with one or more ORT run almost parallel to or even cross the $X = Y$ line, indicating that the tests are useless for pregnancy prediction. AFC, antral follicle count; AMH, anti-Müllerian hormone; FSH, follicle stimulating hormone; ORT, ovarian reserve test; ROC, receiver-operating characteristic.

cut-off levels for poor response prediction could be derived that have a general applicability. Unfortunately, the methods used for assessment of follicle numbers and AMH and FSH serum levels varied across the studies, thereby prohibiting the calculation of relevant cut-off levels. To some extent, correction factors to standardize the results from various studies could be applied. Currently, however, this approach has not yielded final data for one of the three tests of interest. Therefore, centers for ART, applying tests for poor response prediction should rely on their own data analyses for cut-off level assignment. Indeed, development of centre-based prediction models for patient management or counseling is now gaining rapid attention (Banerjee *et al.*, 2010).

Another possible issue may stem from the question of whether the poor responder patient observed in studies is indeed a genuine poor responder or merely the victim of insufficient FSH dosing. Response to ovarian hyperstimulation will mainly depend on the number of follicles sensitive to FSH in a certain time period. With dosages of at least 150 IU, the vast majority of patients will allow all such follicles to develop into dominant follicles. A dose–response effect seems to be only present in dosages under 150 IU daily (Sterrenburg *et al.*, 2011). For the women with a small number of FSH sensitive follicles, applying high dosages, such as 300 or 450 IU would seem ineffective (Harrison *et al.*, 2001; Klinkert *et al.*, 2005a, b; Lekamge *et al.*, 2008), although some studies offer some hope in this respect (Popovic-Todorovic *et al.*, 2003a, b). Indeed, applying massive dosages of FSH may even affect the quality of the oocytes obtained from the responding follicles (Check *et al.*, 2007). However, to evaluate the effect of FSH dosage in the prediction of a poor response and ongoing pregnancy, FSH dosage was added to the multivariable models. Although FSH dosage had a significant role in the models for the prediction of a poor response, it did not alter the predictive capacity of age and ORTs. Importantly a higher starting dosage of FSH was associated with a poor response. So, although FSH dosage affects the prediction of a poor response, the predictive effect may well be based on verification bias and therefore represent a spurious relation. Regarding the identified predictors, ORTs and female age, it can now be underlined that ovarian stimulation has been maximal for the great majority of cases included in this analysis. The occurrence of a poor response cannot therefore be explained by under dosing, but is based on poor follicle number, expressed by the predictive tests female age and ORTs. For ongoing pregnancy prediction, FSH dosage did not have a significant role in the models, neither did it alter the predictive capacity of age and ORTs.

The clinical implications of the present findings will necessarily remain limited to the use of ORTs in predicting poor response to controlled ovarian hyperstimulation for IVF. The real clinical value of the prediction of a poor response will depend on the consequences of the prediction. So far, clinicians do not agree on what alterations in treatment regimen may be of help improving pregnancy prospects in predicted poor responders (Tarlatzis *et al.*, 2003; Shanbhag *et al.*, 2007; Sunkara *et al.*, 2007). Various (pseudo)randomized controlled trials have investigated whether individualization of the FSH treatment dose results, in not necessarily a higher ovarian response but, in higher pregnancy chances in poor responders (Harrison *et al.*, 2001; Popovic-Todorovic *et al.*, 2003a; Klinkert *et al.*, 2005a; Lekamge *et al.*, 2008; Olivennes *et al.*, 2009). Only one study reported a dosing algorithm that would increase pregnancy chances in poor responders, while others were unable to reproduce these effects (Popovic-Todorovic *et al.*, 2003a). For optimization of the ovarian response, two studies have shown that with an individual dose, the response could be optimized and fewer patients would have a poor response (Popovic-

Todorovic et al., 2003a; Olivennes et al., 2009). This could have consequences for the treatment efficacy and costs. Future large, well-designed randomized controlled trials are necessary to identify the best treatment option for poor responders. At present, due to the low accuracy of ORTs in pregnancy prediction, exclusion of patients other than on the basis of female age is not to be supported.

In conclusion, this IPD meta-analysis demonstrates that the ORTs, AFC and AMH are highly capable of forecasting a poor responder to ovarian hyperstimulation for IVF, even without using female age. The clinical applicability of ORT-based dose adaptation on efficacy and costs remains to be demonstrated. Even more importantly, correctly identifying patients with a very poor prognosis for success in ART will not be improved by any currently known ORT. In the field of patient selection prior to ART, female age therefore remains the most important, though modestly effective, tool.

Supplementary data

Supplementary data are available at http://humupd.oxfordjournals.org/.

Authors' roles

R.A.A., M.A., L.B., E.C., A.B.C., T.E., T.E.-G., M.E., E.M.G., K.J., N.R.-F., E.K., J.K., A.L.M., C.B.L., M.M., L.T.M., S.M., S.M.N., H.Y.N., B.P., J.M.J.S., C.T., P.J.Q.V.L., I.K.V. and F.J.M.B. were responsible for data collection. S.L.B., J.D., K.A.B., B.C.O., P.B. and M.J.C.E. performed the analyses. S.L.B., J.D., K.A.B., M.D., B.C.O., P.B., M.J.C.E., B.W.M. and F.J.M.B. made the interpretation of the results and involved in the writing of the article. All authors revised the article.

Funding

This project was funded by CVZ (Assembly of health insurances in the Netherlands). However, no funding bodies had any role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of interests

Prof. F.J.M. Broekmans is a member of the external advisory board for Ferring Pharmaceuticals, Hoofddorp, the Netherlands. He receives no monetary compensation. All other authors have no potential conflict of interests.

Appendix

IMPORT study group – full list of authors:

S.L. Broer, MD, PhD, S.L.Broer-2@umcutrecht.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands; **J. van Disseldorp, MD, PhD**, J.vanDisseldorp-2@umcutrecht.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands; **K.A. Broeze, MD**, K.A.Broeze@amc.uva.nl, Center for Reproductive Medicine, Department of Obstetrics and Gynaecology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; **M. Dolleman, MD**, M.Dolleman-2@umcutrecht.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands; **B.C. Opmeer, PhD**, B.C.Opmeer@amc.uva.nl, Academic Medical Center, Department of Clinical Epidemiology and Biostatistics Netherlands; **R.A. Anderson**, richard.anderson@ed.ac.uk, Centre for Reproductive Biology, Reproductive and Developmental Sciences Edinburgh, United Kingdom; **M. Ashrafi**, ashrafim@royaninstitute.org, Royan Institute Tehran, Islamic Republic of Iran; **L. Bancsi**, L. Bancsi@rijnstate.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands; **E. Caroppo**, ecaroppo@teseo.it, ASL Bari, IVF Unit Bari, Italy; **A. Copperman**, acopperman@rmany.com, Reproductive Medicine Associates of New York New York, United States; **T. Ebner**, thomas.ebner@gespag.at, IVF Unit, Womens' General Hospital Linz Linz, Austria; **M. Eldar Geva**, gevat@szmc.org.il, Shaare-Zedek Medical Center, IVF Unit Israel; **M. Erdem**, erdemom@yahoo.com, Gazi University School of Medicine, Department of Obstetrics and Gynecology Ankara, Turkey; **E.M. Greenblatt**, egreenblatt@mtsinai.on.ca, Mount Sinai Center for Fertility and Reproductive Health, Department of Reproductive Endocrinology and Infertility Toronto, Ontario, Canada; **K. Jayaprakasan**, k.jayaprakasan@nottingham.ac.uk, Academic Division of Reproductive Medicine and Surgery, School of Human Development, University of Nottingham, Nottingham, UK; **Raine Fenning**, Nick.Fenning@nottingham.ac.uk, Academic Division of Reproductive Medicine and Surgery, School of Human Development, University of Nottingham, Nottingham, UK; **E.R. Klinkert**, erklinkert@planet.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands; **J. Kwee**, j.kwee@slaz.nl, VU University Medical Center-Department of Obstetrics/Gynaecology, Amsterdam, The Netherlands; **C.B. Lambalk**, cb.lambalk@vumc.nl, VU University Medical Center Department of Obstetrics/Gynaecology, Amsterdam, The Netherlands; **A. La Marca**, antlamarca@libero.it, University of Modena and Reggio Emilia, Mother-Infant Department, Section of Obstetrics and Gynecology Modena, Italy; **M. McIlveen**, myvanwy@mcilveen.com.au, University of Sydney, Department of Obstetrics and Gynaecology Sydney, New South Wales, Australia; **L.T. Merce**, ltmerce@sego.es, Ruber International Hospital, Department of Obstetrics and Gynaecology Madrid, Spain; **S. Muttukrishna**, s.muttukrishna@ucc.ie, s.muttukrishna@ucl.ac.uk, Royal Free and University College Medical School, Department of Obstetrics and Gynaecology, United Kingdom; **S.M. Nelson**, s.nelson@clinmed.gla.ac.uk, University of Glasgow, Reproductive & Maternal Medicine Glasgow, United Kingdom; **H.Y. Ng**, nghye@hku.hk, University of Hong Kong, Department of Obstetrics and Gynaecology Hong Kong, Hong Kong; **B. Popovic-Todorovic**, drbiba@yahoo.com, Copenhagen University Hospital, Rigshospitalet, The Fertility Clinic Copenhagen, Denmark; **J.M.J. Smeenk**, j.smeenk@elisabeth.nl, University Hospital Nijmegen, Department of Obstetrics and Gynaecology, Nijmegen, Netherlands; **C. Tomás**, ctomas@avaclinic.com, AVA Clinic, Fertility Center Tampere, Finland; **P.J.Q. Van der Linden**, p.j.q.vanderlinden@dz.nl, Deventer Hospital, Department of Obstetrics and Gynaecology Deventer, Netherlands; **I.A. van Rooij**, iajvanrooij@gmail.com, Twee Steden Ziekenhuis, Obstetrics and Gynaecology Tilburg, Netherlands; **I.K. Vladimirov**, vladimirov@doctor.bg, Medical University Sofia, Obstetrics and Gynecology Sofia, Bulgaria; **P.B. Bossuyt**, p.m.bossuyt@amc.uva.nl,

Academic Medical Center, Department of Clinical Epidemiology, Biostatistics and Bioinformatics Amsterdam, Netherlands; **M.J.C. Eijkemans**, M.J.C.Eijkemans@umcutrecht.nl, University Medical Center Utrecht, Julius Center for Health Sciences and Primary care Utrecht, Netherlands; **B.W.Mol**, B.w.mol@amc.uva.nl, Academic Medical Center, Obstetrics and Gynaecology Amsterdam, Netherlands; **Frank Broekmans**, F.Broekmans@umcutrecht.nl, University Medical Center Utrecht, Division Woman and Baby, Department or Reproductive Medicine, Utrecht, Netherlands.

References

Ashrafi M, Madani T, Tehranian AS, Malekzadeh F. Follicle stimulating hormone as a predictor of ovarian response in women undergoing controlled ovarian hyperstimulation for IVF. *Int J Gynaecol Obstet* 2005;**91**:53–57.

Bancsi LF, Huijs AM, Den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, te Velde ER. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. *Fertil Steril* 2000; **73**:552–557.

Banerjee P, Choi B, Shahine LK, Jun SH, O'Leary K, Lathi RB, Westphal LM, Wong WH, Yao MW. Deep phenotyping to predict live birth outcomes in in vitro fertilization. *Proc Natl Acad Sci USA* 2010;**107**:13570–13575.

Barad DH, Weghofer A, Gleicher N. Age-specific levels for basal follicle-stimulating hormone assessment of ovarian function. *Obstet Gynecol* 2007;**109**:1404–1410.

Boomsma CM, Macklon NS. What can the clinician do to improve implantation? *Reprod Biomed Online* 2006;**13**:845–855.

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.

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.

Broer SL, Eijkemans MJ, Scheffer GJ, van Roja IA, de Vet A, Themmen AP, Laven JS, de Jong FH, te Velde ER, Fauser BC *et al.* Anti-mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. *J Clin Endocrinol Metab* 2011;**96**:2532–2539.

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.

Broeze KA, Opmeer BC, van der Veen F, Bossuyt PM, Bhattacharya S, Mol BW. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. *Hum Reprod Update* 2010;**16**:561–567.

Caroppo E, Matteo M, Schonauer LM, Vizziello G, Pasquadibisceglie A, Vitti A, D'Amato G. Basal FSH concentration as a predictor of IVF outcome in older women undergoing stimulation with GnRH antagonist. *Reprod Biomed Online* 2006;**13**:815–820.

Check JH, Summers-Chase D, Yuan W, Horwath D, Wilson C. Effect of embryo quality on pregnancy outcome following single embryo transfer in women with a diminished egg reserve. *Fertil Steril* 2007;**87**:749–756.

de Vet A, Laven JSE, de Jong FH, Themmen A, Fauser BCJM. Anti-Mullerian Hormone serum levels: A putative marker for ovarian aging. *Fertil Steril* 2002; **77**:357–362.

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.

Erdem M, Erdem A, Gursoy R, Biberoglu K. Comparison of basal and clomiphene citrate induced FSH and inhibin B, ovarian volume and antral follicle counts as ovarian reserve tests and predictors of poor ovarian response in IVF. *J Assist Reprod Genet* 2004;**21**:37–45.

Harrison RF, Jacob S, Spillane H, Mallon E, Hennelly B. A prospective randomized clinical trial of differing starter doses of recombinant follicle-stimulating hormone (follitropin-beta) for first time in vitro fertilization and intracytoplasmic sperm injection treatment cycles. *Fertil Steril* 2001;**75**:23–31.

Hendriks DJ, te Velde ER, Looman CW, Bancsi LF, Broekmans FJ. The role of poor response in the prediction of the cumulative ongoing pregnancy rate in in vitro fertilisation. In: *Dynamic and Basal Ovarian Reserve Tests for Outcome Prediction in IVF: Comparisons and Meta-Analyses*. Academic Thesis. Utrecht 2005;162–179.

Henne MB, Stegmann BJ, Neithardt AB, Catherino WH, Armstrong AY, Kao TC, Segars JH. The combined effect of age and basal follicle-stimulating hormone on the cost of a live birth at assisted reproductive technology. *Fertil Steril* 2008; **89**:104–110.

Janes H, Longton G, Pepe MS. Accommodating covariates in receiver operating characteristic analysis. *Stata J* 2009;**9**:17–39.

Jayaprakasan K, Hilwah N, Kendall NR, Hopkisson JF, Campbell BK, Johnson IR, Raine-Fenning NJ. Does 3D ultrasound offer any advantage in the pretreatment assessment of ovarian reserve and prediction of outcome after assisted reproduction treatment? *Hum Reprod* 2007;**22**:1932–1941.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. 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* 2005a;**20**:611–615.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. 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* 2005b;**83**:811–814.

Kwee J, Elting MW, Schats R, Bezemer PD, Lambalk CB, Schoemaker J. Comparison of endocrine tests with respect to their predictive value on the outcome of ovarian hyperstimulation in IVF treatment: results of a prospective randomized study. *Hum Reprod* 2003;**18**:1422–1427.

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, Nelson SM, Sighinolfi G, Manno M, Baraldi E, Roli L, Xella S, Marsella T, Tagliasacchi D, D'Amico R *et al.* Anti-Mullerian hormone-based prediction model for a live birth in assisted reproduction. *Reprod Biomed Online* 2011; **22**:341–349.

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;**149**:889–897.

Lekamge DN, Lane M, Gilchrist RB, Tremellen KP. Increased gonadotrophin stimulation does not improve IVF outcomes in patients with predicted poor ovarian reserve. *J Assist Reprod Genet* 2008;**25**:515–521.

Lintsen AM, Eijkemans MJ, Hunault CC, Bouwmans CA, Hakkaart L, Habbema JD, Braat DD. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Hum Reprod* 2007;**22**:2455–2462.

Liu KE, Greenblatt EM. Elevated day 3 follicle-stimulating hormone/luteinizing hormone ratio $>$ or $=2$ is associated with higher rates of cancellation in in vitro fertilization-embryo transfer cycles. *Fertil Steril* 2008;**90**:297–301.

Luna M, Grunfeld L, Mukherjee T, Sandler B, Copperman AB. Moderately elevated levels of basal follicle-stimulating hormone in young patients predict low ovarian response, but should not be used to disqualify patients from attempting in vitro fertilization. *Fertil Steril* 2007;**87**:782–787.

Malizia BA, Hacker MR, Penzias AS. Cumulative live-birth rates after in vitro fertilization. *N Engl J Med* 2009;**360**:236–243.

McIlveen M, Skull JD, Ledger WL. Evaluation of the utility of multiple endocrine and ultrasound measures of ovarian reserve in the prediction of cycle cancellation in a high-risk IVF population. *Hum Reprod* 2007;**22**:778–785.

Merce LT, Barco MJ, Bau S, Troyano JM. Prediction of ovarian response and IVF/ICSI outcome by three-dimensional ultrasonography and power Doppler angiography. *Eur J Obstet Gynecol Reprod Biol* 2007;**132**:93–100.

Muasher SJ, Oehninger S, Simonetti S, Matta J, Ellis LM, Liu HC, Jones GS, Rosenwaks Z. The value of basal and/or stimulated serum gonadotropin levels in prediction of stimulation response and in vitro fertilization outcome. *Fertil Steril* 1988;**50**:298–307.

Muttukrishna S, Suharjono H, McGarrigle H, Sathanandan M. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? *BJOG* 2004;**111**:1248–1253.

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

Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. *PLoS Med* 2011;**8**:e1000386.

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.

Ng EH, Chan CC, Tang OS, Ho PC. Antral follicle count and FSH concentration after clomiphene citrate challenge test in the prediction of ovarian response during IVF treatment. *Hum Reprod* 2005;**20**:1647–1654.

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.

Pepe MS, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J* 2009;**9**:1–16.

Popovic-Todorovic B, Loft A, Bredkjaeer HE, Bangsboll S, Nielsen IK, Andersen AN. A prospective randomized clinical trial comparing an individual dose of recombinant FSH based on predictive factors versus a 'standard' dose of 150 IU/day in 'standard' patients undergoing IVF/ICSI treatment. *Hum Reprod* 2003a;**18**:2275–2282.

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* 2003b;**18**:781–787.

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

Scott RT Jr, Elkind-Hirsch KE, Styne-Gross A, Miller KA, Frattarelli JL. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. *Fertil Steril* 2008;**89**:868–878.

Shanbhag S, Aucott L, Bhattacharya S, Hamilton MA, McTavish AR. Interventions for 'poor responders' to controlled ovarian hyperstimulation (COH) in in-vitro fertilisation (IVF). *Cochrane Database Syst Rev* 2007;CD004379.

Smeenk JM, Stolwijk AM, Kremer JA, Braat DD. External validation of the templeton model for predicting success after IVF. *Hum Reprod* 2000;**15**:1065–1068.

Smeenk JM, Sweep FC, Zielhuis GA, Kremer JA, Thomas CM, Braat DD. Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. *Fertil Steril* 2007;**87**:223–226.

Sowers MR, Eyvazzadeh AD, McConnell D, Yosef M, Jannausch ML, Zhang D, Harlow S, Randolph JF Jr. Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. *J Clin Endocrinol Metab* 2008;**93**:3478–3483.

Sterrenburg MD, Veltman-Verhulst SM, Eijkemans MJ, Hughes EG, Macklon NS, Broekmans FJ, Fauser BC. Clinical outcomes in relation to the daily dose of recombinant follicle-stimulating hormone for ovarian stimulation in in vitro fertilization in presumed normal responders younger than 39 years: a meta-analysis. *Hum Reprod Update* 2011;**17**:184–196.

Sunkara SK, Tuthill J, Khairy M, El-Toukhy T, Coomarasamy A, Khalaf Y, Braude P. Pituitary suppression regimens in poor responders undergoing IVF treatment: a systematic review and meta-analysis. *Reprod Biomed Online* 2007;**15**:539–546.

Tarlatzis BC, Zepiridis L, Grimbizis G, Bontis J. Clinical management of low ovarian response to stimulation for IVF: a systematic review. *Hum Reprod Update* 2003; **9**:61–76.

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

van Disseldorp J, Faddy MJ, Themmen AP, de Jong FH, Peeters PH, van der Schouw YT, Broekmans FJ. Relationship of serum anti-Mullerian hormone concentration to age of menopause. *J Clin Endocrinol Metab* 2008;**93**:2129–2134.

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.

van Rooij IA, Tonkelaar I, Broekmans FJ, Looman CW, Scheffer GJ, de Jong FH, Themmen AP, te Velde ER. Anti-mullerian hormone is a promising predictor for the occurrence of the menopausal transition. *Menopause* 2004;**11**:601–606.

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

van Swieten EC, Leeuw-Harmsen L, Badings EA, van der Linden PJ. Obesity and Clomiphene Challenge Test as predictors of outcome of in vitro ferti-lization and intracytoplasmic sperm injection. *Gynecol Obstet Invest* 2005;**59**:220–224.

Verhagen TE, Hendriks DJ, Bancsi LF, Mol BW, Broekmans FJ. The accuracy of multivariate models predicting ovarian reserve and pregnancy after in vitro fertilization: a meta-analysis. *Hum Reprod Update* 2008;**14**:95–100.

Vladimirov IK, Tacheva DM, Kalinov KB. Mean ovarian diameter (MOD) as a predictor of poor ovarian response. *J Assist Reprod Genet* 2004;**21**:73–77.

Vladimirov IK, Tacheva DM, Kalinov KB, Ivanova AV, Blagoeva VD. Prognostic value of some ovarian reserve tests in poor responders. *Arch Gynecol Obstet* 2005; **272**:74–79.

Yong PY, Baird DT, Thong KJ, McNeilly AS, Anderson RA. Prospective analysis of the relationships between the ovarian follicle cohort and basal FSH concentration, the inhibin response to exogenous FSH and ovarian follicle number at different stages of the normal menstrual cycle and after pituitary down-regulation. *Hum Reprod* 2003;**18**:35–44.