ORIGINAL ARTICLE: ASSISTED REPRODUCTION

Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis

Simone L. Broer, M.D., Ph.D.,[a] Madeleine Dólleman, M.D.,[a] Jeroen van Disseldorp, M.D., Ph.D.,[a] Kimiko A. Broeze, M.D.,[b] Brent C. Opmeer, Ph.D.,[c] Patrick M. M. Bossuyt, Ph.D.,[c] Martinus J. C. Eijkemans, Ph.D.,[d] Ben Willem Mol, M.D., Ph.D.,[b] and Frank J. M. Broekmans, M.D., Ph.D.,[a] on behalf of the IPD-EXPORT Study Group

[a] Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht; [b] Department of Obstetrics and Gynaecology, Academic Medical Centre, Amsterdam; [c] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam; and [d] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

Objective: To evaluate whether ovarian reserve tests (ORTs) add prognostic value to patient characteristics, such as female age, in the prediction of excessive response to ovarian hyperstimulation in patients undergoing IVF, and whether their performance differs across clinical subgroups.
Design: Authors of studies reporting on basal FSH, antimüllerian hormone (AMH), or antral follicle count (AFC) in relation to ovarian response to ovarian hyperstimulation were invited to share original data. Random intercept logistic regression models were used to estimate added value of ORTs on patient characteristics, while accounting for between-study heterogeneity. Receiver operating characteristic regression analyses were performed to study the effect of patient characteristics on ORT accuracy.
Setting: In vitro fertilization clinics.
Patient(s): A total of 4,786 women for the main analysis, with a subgroup of 1,023 women with information on all three ORTs.
Intervention(s): None.
Main Outcome Measure(s): Excessive response prediction.
Result(s): We included 57 studies reporting on 32 databases. Female age had an area under the receiver operating characteristic curve of 0.61 for excessive response prediction. Antral follicle count and AMH significantly added prognostic value to this. A model with female age, AFC, and AMH had an area under the receiver operating characteristic curve of 0.85. The combination of AMH and AFC, without age, had similar accuracy. Subgroup analysis indicated that FSH performed significantly worse in predicting excessive response in higher age groups, AFC did significantly better, and AMH performed the same.

Received January 17, 2013; revised and accepted April 12, 2013.
The EXPORT (Excessive Response Prediction using Ovarian Reserve Tests) Study Group: S. L. Broer, M. Dólleman, J. van Disseldorp, K. A. Broeze, B. C. Opmeer, P. M. M. Bossuyt, M. J. C. Eijkemans, B. W. Mol, F. J. M. Broekmans, A. Aflatoonian, R. A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A. B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, T. Freour, C. Gnoth, E. M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, C. B. Lambalk, M. McIlveen, L. Mohiyiddeen, L. T. Merce, S. Muttukrishna, L. G. Nardo, S. M. Nelson, H. Y. Ng, B. Popovic-Todorovic, J. M. J. Smeenk, C. Tomás, P. J. Q. Van der Linden, I. A. van Rooij, and I. K. Vladimirov.
S.L.B. has received payments from the European Society of Human Reproduction and Embryology for travel expenses and as an invited speaker. M.D. has nothing to disclose. J.v.D. has nothing to disclose. K.A.B. has nothing to disclose. B.C.O. has nothing to disclose. P.M.M.B. has nothing to disclose. M.J.C.E. has nothing to disclose. B.W.M. has nothing to disclose. F.J.M.B. is a member of the external advisory board for Ferring Pharmaceuticals, Hoofddorp, The Netherlands, for which he receives no monetary compensation, and has received fees and grant support from Ferring, Gedeon Richter, Merck Serono, MSD, and Roche.
S.L.B. and M.D. contributed equally to this work.
This project was funded by CVZ (Assembly of Health Insurances in The Netherlands). However, no funding bodies had any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Reprint requests: Madeleine Dólleman, M.D., University Medical Center Utrecht, Reproductive Medicine and Gynecology, Room 05.126, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands (E-mail: M.Dolleman-2@umcutrecht.nl).

Fertility and Sterility® Vol. ■, No. ■, ■ 2013 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2013.04.024

Conclusion(s): We demonstrate that AFC and AMH add value to female age in the prediction of excessive response and that, for AFC and FSH, the discriminatory performance is affected by female age. (Fertil Steril® 2013;■:■–■. ©2013 by American Society for Reproductive Medicine.)

Key Words: Assisted Reproduction, antral folicle count, antimullerian horme, excessive response, ovarian hyperstimulation syndrome

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/dollemanm-ivf-ovarian-reserve-amh-ohss/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

In women undergoing IVF, the development of a large number of oocytes occurs in one third of IVF cycles (1, 2). Such an excessive response may lead to poorer-quality embryos, lower chances of pregnancy, or cycle cancellation (3–9). Additionally, patients with an excessive response are at risk of developing ovarian hyperstimulation syndrome (OHSS), a potentially life threatening condition (10–12). To maximize safety and efficacy of assisted reproductive technology programs, there is a need to identify patients at risk of an excessive response at the start of IVF/intracytoplasmic sperm injection and to apply effective measures to prevent such an excessive response from occurring.

Several patient characteristics, such as a lean habitus, young age, and the presence of polycystic ovary syndrome (PCOS), have been identified as conditions that predispose patients to OHSS (13). Unfortunately, precise expressions of the predictive accuracy of these characteristics are not available. In contrast, ovarian reserve tests (ORTs), such as antimüllerian hormone (AMH), antral follicle count (AFC), and FSH, have been assessed for their value in the prediction of an excessive response (4, 6, 14–27). It is not clear, however, whether these ORTs add to predictive and readily available patient characteristics, of which female age is the most important.

Because ovarian reserve decreases with increasing age, it is conceivable that the predictive value of the ORTs is mutually dependent on female age. Alternatively, the accuracy of AFC may be different in women with a higher body mass index (BMI). Moreover, BMI could further influence the predictive accuracy by possibly reducing the biologic availability of recombinant FSH for ovarian stimulation, thereby creating spuriously reduced ovarian responses (28). Most predictive accuracy studies, however, had a limited sample size, lacking the power to evaluate patient characteristics as modifiers of accuracy in specific subgroups and the ability to analyze the added value of the ORTs on patient characteristics.

To overcome the problem of small studies with restricted power, the present study applied an individual patient data (IPD) meta-analysis approach. By aggregating data on the level of the individual patient, more precise estimates of accuracy, evaluations of added accuracy, and identification of accuracy modifiers become possible, while taking between-study heterogeneity into account appropriately.

MATERIALS AND METHODS

Data Acquisition

We searched the existing literature for studies on the value of FSH, AFC, and AMH in predicting IVF outcome. We expanded searches from conventional systematic reviews on the subject and another IPD meta-analysis (IPD-Individual patient data Meta-analysis on Poor response prediction with Ovarian Reserve Tests [IMPORT]) on poor response prediction; searches were updated to include studies up to the end of 2009 (14, 29–32).

Key words used in the systematic MEDLINE search included synonyms for in vitro fertilization (IVF, controlled ovarian stimulation, in vitro fertilisation) and synonyms for the various tests (FSH, follicle stimulating hormone, AFC, antral follicle count or number, AMH, antimüllerian hormone, müllerian inhibiting substance). Studies presenting data on ovarian response to hyperstimulation, at least one ORT, and at least one patient characteristic were eligible for the present review. All titles and abstracts were evaluated for eligibility by two authors (M.D. and S.L.B. or S.L.B. and J.v.D.). If necessary, the opinion of a third author was decisive (F.J.M.B.).

All authors of potentially eligible primary studies were informed about this IPD meta-analysis initiative 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 on the format of the data requested.

After data acquisition, all data were scrutinized on quality and consistency and, whenever possible, converted into a single format. Any issues or inconsistencies were checked with the original author. For a more detailed description of the IPD meta-analysis methodology the reader is referred to previous articles (33, 34).

Within all eligible studies, a comparison was made between those studies that could and those that could not be included. Sensitivity–specificity pairs for excessive response prediction were calculated for the ORTs under study, using the thresholds for excessive response that had been set in each study. Spearman correlations were then calculated for sensitivity–specificity pairs across studies, to ascertain that the differences in sensitivity and specificity levels between included and not-included studies were likely the result of different threshold levels used, thereby reducing the likelihood of bias in the final analysis.

All original studies either had approval of their local research ethics committee or were exempt from obtaining such approval

owing to the nature of the study. We evaluated the quality of the included studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist, supplemented by a number of items to evaluate the risk of bias in prognostic studies. Whenever a particular variable was missing in an individual database or in an individual case within a database, data were not imputed. Baseline characteristics were analyzed in the total IPD dataset and for each of the individual studies.

Definitions

An excessive response was defined as the retrieval of more than 15 oocytes. This cutoff was selected because the definition for excessive responsive in most primary studies varied between more than 14 and more than 16 oocytes (6, 17, 19, 23, 35–38). Furthermore, it has been shown that clinical pregnancy rates decline with the retrieval of more than 15 oocytes, thus arguing that it is an unfavorable condition (39). Duration of subfertility was defined as the period from cessation of oral contraceptives and/or start of unprotected intercourse until the first IVF attempt. In the included studies, patients had been stimulated according to local protocol, resulting in a wide range of daily FSH dosages. In almost all studies a starting dosage of at least 150 IU was given. This dosage is considered the optimal daily dosage in expected normal responders; with this dose it may be assumed that all patients received adequate stimulation, creating growth of all follicles sensitive to FSH within the time frame of exposure (40).

Predictive accuracy was defined as the ability of the model to distinguish excessive responders from cases with a normal or poor response. We calculated areas under the receiver operating characteristic (ROC) curve (AUCs) for the ORTs in the prediction of excessive response for each individual study and for the pooled studies, calculated as a summary statistic of predictive accuracy.

Statistical Analysis

Analyses were done in two steps. First, the added value of ORTs on top of the patient characteristics age, BMI, and duration of subfertility was assessed. As a part of this analysis, we assessed whether these results may have been influenced by differences in study characteristics or daily FSH dosage administered. Second, we examined whether the predictive performance depends on the patient characteristics age, BMI, and duration of subfertility.

Prediction of an excessive response using ORTs and patient characteristics. To study whether ORTs have an added value on top of patient characteristics in the prediction of an excessive response, we used random intercept logistic regression models. 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 incidence of excessive response in this universe can be described by a normal distribution on the log odds scale. These models were created to quantitatively estimate the added value that ORTs have on patient characteristics in predicting an excessive response. It provides both an estimate of the summary predictive effect as well as of the variance of the between-study distribution of the incidence of excessive response.

Three different sets of models were used for the prediction of excessive response. The first set of models included the patient characteristics female age, BMI, and duration of subfertility. In the second set of models, the predictive capacity of each of the individual ovarian reserve tests (FSH, AFC, and AMH) was estimated. In the third set of multivariate models, the added value of combinations of ovarian reserve tests on top of patient characteristics was evaluated.

The next step was to construct ROC curves to express the predictive accuracy of each combination of predictive variables in distinguishing excessive responders from the rest. With each of the random intercept logistic regression models, we calculated the probability of an excessive response. By moving the positivity threshold from 0 to 1, we could then calculate sensitivity–specificity pairs for each model. On the basis of these, we plotted stratified ROC curves with the ROC regression model as proposed by Janes, Pepe, and colleagues (41, 42). This model assumes that studies share a common ROC for each ORT 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 AUCs of the models described above and evaluate the statistical significance of any differences.

Because not all studies in this meta-analysis had included data for all three ORTs, we constructed prediction models using those databases from the total dataset that included the corresponding ovarian reserve tests (FSH, AFC, and AMH) and age to allow for a direct comparison. The results of all analyses in the three-test study subgroup were verified in the total study group.

Because it was not recorded whether studies adjusted FSH dosage according to results of the ovarian reserve tests and because this may have been different between fertility physicians, correction on the level of the individual study was not considered to be enough, and correction on the individual level was necessary.Therefore, we repeated the analyses as described above while adding starting daily FSH dosage as a covariate. In a similar fashion, we included study design features, as identified by the QUADAS checklist, as covariates in our models, to evaluate whether differences in daily FSH dosage or study design influenced the observed associations between ORT, patient characteristics, and the outcome excessive response (43).

Influence of age, BMI, and duration of subfertility on the accuracy of ORTs in excessive response prediction. To study whether the accuracy of ORTs in the prediction of excessive response is modified by patient age, BMI, or duration of subfertility, we used the ROC regression model proposed by Pepe, Janes, and colleagues (41, 42). This model allows us to study the effects of patient or disease characteristics on the classification accuracy of tests. In this model, the ORT ROC curves are modeled as a function of the covariates age, BMI, and duration of subfertility.

We assumed the effect of the covariate in this meta-analysis to be identical across studies, but, as in the previous analysis, the positivity threshold corresponding to each sensitivity–specificity pair was allowed to vary between studies, thereby correcting for any heterogeneity between studies. The areas under the corresponding ROC curves were calculated, to express the discriminatory capacity (accuracy) of the ORT in women in the respective subgroups.

Data were analyzed using SPSS 17.0 (SPSS Inc.) and R version 2.9.0 (http://www.r-project.org/). Random intercept logistic regression prediction models were created with the 'Lme4' library, using the Laplace approximation to the likelihood.

RESULTS

Data Acquisition

The MEDLINE search up to the end of 2009 delivered 2,551 hits, of which 125 were eligible for inclusion. In 22 studies the authors were untraceable, 33 authors did not reply after repeated effort, in 12 studies the data were lost, and 2 studies were not suited for the current analysis. This resulted in a total of 32 databases, used for the preparation of 57 or more manuscripts, which could be included in this IPD study. Twenty-seven had been previously included in the IPD-IMPORT study (32). Ten additional studies were identified from the systematic MEDLINE search. We invited these 10 extra authors and asked them, as well as the previous 27 studies, for permission to use their databases in the present analysis on excessive response prediction. Only four of these authors sent their data (11, 12, 25, 37), one of them submitting two separate databases (25). In total, 32 datasets could be included in the EXPORT study project database, with data from 5,251 study participants (Supplemental Fig. 1, available online) (2, 6, 9, 11, 12, 15–19, 21, 24–27, 36–38, 44–56).

We were able to replicate the primary findings of the original study in 13 databases. In 12 cases, the study database we received contained a number of patients that differed from the publication, whereas in seven other databases there were slight inconsistencies with the baseline data as previously published. These inconsistencies were discussed with the corresponding author and could be resolved in all cases. Through this process, the level of consistency between the individual data and the data reported in the published articles was regarded sufficient for all included studies.

For the comparison of the four included and the six not-included studies, we attempted to calculate sensitivity and specificity of the ORTs in the prediction of excessive response. However, of the nonincluded studies only one reported sensitivity and specificity values for AFC in the prediction of an excessive response (23). Therefore, Spearman correlation could not be calculated. Nonetheless, for the majority of the studies this was performed in the IMPORT study (32), a related IPD study from the same research group focused on poor response prediction. In that study it was demonstrated that there was no difference in the correlations between sensitivity and specificity for included and not-included studies on poor response. Because there was no difference in poor response prediction, it is reasonable to assume that there is also no difference for excessive response prediction. We therefore assumed that no obvious bias has occurred for the present analysis by excluding studies on the basis of the availability of primary data. Baseline characteristics of the original studies are summarized in Supplemental Figure 3.

Data from 4,786 of the 5,251 women were suitable for the analysis of prediction of excessive response, of which 894 (19%) had an excessive response. In the other 465 women information on oocyte yield was missing. Baseline characteristics of the total study group are summarized in Table 1. The AUCs of the original studies for excessive response prediction are summarized in Supplemental Table 1.

Statistical Analyses

Prediction of an excessive response using ORTs and patient characteristics. For the model building exercises, we could use data of 1,023 women from 10 datasets for excessive response analysis. This was the number of women for whom all five variables of interest were known: age, AFC, AMH, FSH, and the number of oocytes retrieved after stimulation. Of the evaluated patient characteristics, age was the strongest single predictor of excessive response (odds ratio [OR] 0.89, 95% confidence interval [CI] 0.85–0.93), as shown in Table 2. Body mass index and duration of subfertility were not significantly predictive of excessive response (Supplemental Table 2).

We compared the ORTs using the random intercept logistic regression model in predicting excessive response. The ROC regression analysis showed a high accuracy for

TABLE 1

Baseline characteristics from pooled data.

Characteristic	Total population (n = 4,786)	Excessive responders (n = 894)	Nonexcessive responder (n = 3,892)	*P* value
Female age (y)	34.4 (26.0–42.0)	32.5 (25.0–39.9)	34.7 (26.0–42.0)	< .001
FSH (IU/L)	7.7 (3.8–14.0)	6.4 (3.5–10.1)	8.7 (3.9–16.0)	< .001
AFC (n)	12.1 (3.0–25.6)	17.1 (6.0–32.0)	11.0 (3.0–22.0)	< .001
AMH (ng/mL)	2.5 (0.1–7.6)	4.8 (1.3–10.2)	2.0 (0.1–5.7)	< .001
BMI (kg/m^2)	23.6 (18.6–30.1)	23.4 (18.5–29.4)	23.4 (18.6–30.1)	.943
Duration of subfertility (y)	4.3 (1.3–10.0)	4.3 (1.5–10.0)	4.3 (1.2–10.0)	.937

Note: Excessive response definition: >15 oocytes retrieved. Duration of subfertility: the period from the cessation of contraceptive methods or start of unprotected intercourse until the first IVF attempt. Values are presented as mean (5th percentile–95th percentile).

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

TABLE 2

Univariable and multivariable models of age and ORTs in the prediction of an excessive response.

	Three-test study group (n = 1,023)				Total study group (n = 4,786)			
Models	**OR**	**95% CI**	***P* value**	**Variance RI**	**OR**	**95% CI**	***P* value**	**Variance RI**
Univariable models								
Age (per year)	0.89	0.85–0.93	<.001	0.748	0.90	0.88–0.91	<.001	0.543
FSH (per IU/L)	0.76	0.70–0.84	<.001	1.23	0.83	0.80–0.86	<.001	0.551
AFC (per no.)	1.18	1.15–1.22	<.001	0.715	1.14	1.12–1.16	<.001	0.605
AMH (per 1.0 ng/mL)	1.61	1.48–1.76	<.001	0.878	1.59	1.49–1.70	<.001	0.680
Multivariable models								
Age and FSH								
Age (per year)	0.91	0.87–0.94	<.001	0.82	0.91	0.89–0.93	<.001	0.497
FSH (per IU/L)	0.79	0.72–0.87	<.001		0.85	0.82–0.88	<.001	
Age and AFC								
Age (per year)	0.93	0.89–1.98	.003	0.769	0.95	0.92–0.98	.001	0.575
AFC (per no.)	1.17	1.13–1.21	<.001		1.13	1.11–1.15	<.001	
Age and AMH								
Age (per year)	0.92	0.88–0.97	<.001	0.596	0.92	0.89–0.95	<.001	0.599
AMH (per 1.0 ng/mL)	1.57	1.43–1.71	<.001		1.54	1.44–1.64	<.001	

Note: Results of random intercept logistic regression models in the prediction of an excessive response. Multivariable analyses showed that all three ORTs add predictive information to female age alone. *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 SD and may be interpreted on the logistic scale. A 1-SD difference between two studies in the population of studies corresponds to an increase in the odds on the outcome (excessive response) of exp(SD). For example, the Age and AMH model for excessive response has Variance RI = 0.321, so exp(sqrt(0.321)) = 1.76 is the relative increase in odds of excessive response, corresponding to a difference between two studies in intercept of 1 SD.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

AMH (AUC 0.81, 95% CI 0.76–0.87) and for AFC (AUC 0.79, 95% CI 0.74–0.84), but only a moderate accuracy for FSH (AUC 0.66, 95% CI 0.60–0.73) (Fig. 1A).

The multivariable analyses demonstrated that a model including age, AFC, and AMH (AUC 0.85) had a significantly higher predictive accuracy than a model based on age alone (AUC 0.61; $P<.001$). Addition of FSH to this model did not further improve predictive accuracy (AUC 0.85; $P=.73$) (Fig. 1A). Interestingly, a single AMH or AFC test had a comparable accuracy (AUC 0.81 and 0.79, respectively). Addition of AMH to AFC and of AFC to AMH significantly improved accuracy ($P<.001$ or $P=.003$, respectively). A model combining these two tests resulted in an AUC of 0.85. Age did not add value to this model ($P=.98$). The ROC curves corresponding to the multivariable models are shown in Figure 1B.

Effect of daily FSH dosage and study protocol on excessive response outcome. Patients had been stimulated with a wide range of daily FSH dosages, according to their center's local protocol. The mean daily FSH dosage was 204.28 IU (interquartile range 150–225 IU). Twenty-one women received daily FSH dosages <150 IU because of an expected excessive response (5 women received 75 IU, 14 women received 112.5 IU, and 2 women received 125 IU of daily FSH). Women who developed an excessive response tended to have received a lower starting dosage of FSH than women who did not develop an excessive response. The mean dosage was 201.75 IU in those women who developed an excessive response, vs. a mean dosage of 224.79 IU for women who did not have an excessive response (*P* value for difference <.001). Daily FSH dosage had a significant, negative association with excessive response development. A higher daily FSH dosage was associated with a lower chance of an excessive response in both the three-test study group and in the group as a whole (OR 0.99, $P<.001$). In the individual studies it was often not stated whether daily FSH dosage protocols were altered according to the results of the ORTs that were measured. Because it is very likely that this occurred and because it is further likely that different physicians acted differently to ORT results, adjusting at the level of the individual study was deemed to not be enough, and correction on the individual level was necessary. When daily FSH dosage was included in the multivariable model as an additional covariate (in addition to age and the ORTs), the ORs for age and the ORTs, adjusted for FSH dosage, remained basically unchanged. In the multivariable model for age, FSH, and daily FSH dosage, FSH had an OR of 0.86 (95% CI 0.83–0.90), the OR for AFC in the multivariable model for age, AFC, and daily FSH dosage remained 1.13 (95% CI 1.11–1.15), and the OR for AMH in the multivariable for age, AMH, and FSH dosage was 1.55 (95% CI 1.45–1.66).

Study quality characteristics as scored by QUADAS checklist and supplemental questions are shown in Supplemental Figure 3. Overall, data were of high quality, with the exception of verification bias. This implies that the test results may have been known to the clinician taking decisions on patient management. None of the study characteristics that were assessed were associated with excessive response development (*P* value range .34–.89). Similarly, the ORs for age and the ORTs, adjusted for study characteristics, remained basically unchanged.

Influence of age, BMI, and duration of subfertility on the accuracy of ORTs in excessive response prediction. The results of the ROC regression model that studied the effect of several patient characteristics on the ROC curve of the ORTs in the prediction of an excessive response are shown in Table 3. The accuracy of FSH was significantly lower in women with a higher age ($P=.01$).

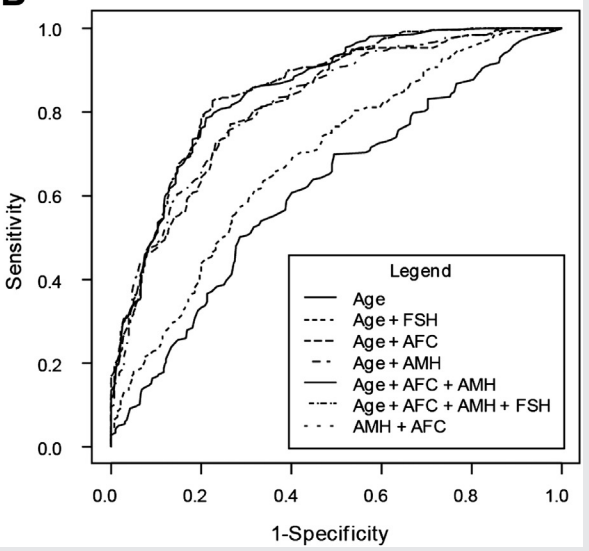
For a 20 year old the AUC for FSH was 0.66. In contrast, the AUC for a 30 year old was 0.59 and for a 40 year old was

FIGURE 1

A

	Three-test study group (N= 1,023)				Total study group (N=4,786)			
	AUC	95% CI	P value	N	AUC	95% CI	P value	N
Univariable analysis								
Age	0.61	0.54 - 0.68	NA	1023	0.61	0.58 - 0.64	NA	4650
FSH	0.66	0.60 - 0.73	0.071	1023	0.64	0.61 - 0.67	0.026	4254
AFC	0.79	0.74 - 0.85	< 0.001	1023	0.73	0.69 - 0.77	< 0.001	2524
AMH	0.81	0.76 - 0.87	< 0.001	1023	0.82	0.77 - 0.86	< 0.001	1890
Multivariable analysis								
Age & FSH	0.68	0.62 - 0.75	< 0.001	1023	0.67	0.64 - 0.71	< 0.001	4254
Age & AFC	0.81	0.76 - 0.87	< 0.001	1023	0.75	0.71 - 0.79	< 0.001	2524
Age & AMH	0.81	0.76 - 0.87	< 0.001	1023	0.81	0.77 - 0.85	< 0.001	1890
Age & AMH & AFC	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1024
Age & AMH & AFC & FSH	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1023
AMH & AFC	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1024

B



Areas under the curve and ROC curves of prediction models of age and ovarian reserve tests for the prediction of an excessive response. (A) Areas under the curves of prediction models of age or ORTs in the prediction of an excessive response are shown. In the univariable analysis it is shown that both AMH and AFC have high accuracy, whereas FSH only has 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 with the model based on age alone. Adding any of the ORTs shows 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. Addition of FSH to this model did not improve the predictive accuracy (*P*=.725). However, a model with AMH and AFC alone has a comparable AUC. (B) Receiver operating characteristic curves of age and ORTs in the prediction of an excessive response. The ROC curves of age and age combined with a single or more ORTs are depicted. The ROC curves for Age + AMH, Age + AFC, Age + AMH + AFC, and Age + AMH + AFC + FSH run toward the upper left corner of the ROC space, indicating a good capacity to discriminate between normal and excessive responders at certain cutoff levels. Receiver operating characteristic curves in the three-test study group (n = 1,023).

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

0.52. The accuracy of AFC was significantly higher in women with a higher age (*P*=.01). For a 20-year-old woman the AUC for AFC was 0.64, for a 30 year old it was 0.71, and for a 40 year old it was 0.81. The discriminatory capacity of AMH in response prediction was not significantly influenced by age. Body mass index and duration of subfertility categories had no significant effect on the ROC curves, for any of the ORTs.

TABLE 3

Results of the ROC regression analysis.

Variable	Coefficient	95% CI	*P* value
Age			
FSH	−0.029	−0.051, −0.006	.010[a]
AFC	0.032	0.006, 0.056	.010[a]
AMH	−0.021	−0.049, 0.005	.139
BMI			
FSH	0.026	−0.024, 0.070	.267
AFC	−0.009	−0.048, 0.033	.674
AMH	0.019	−0.024, 0.056	.363
Duration			
FSH	0.018	−0.044, 0.078	.569
AFC	0.047	−0.022, 0.112	.177
AMH	−0.041	−0.113, 0.026	.246

Note: ROC regression analysis showing the effect of the patient characteristics on the ROC curve of the ORTs in the prediction of an excessive ovarian response. Duration = duration of subfertility.

[a] Significant influence of the patient characteristics on the discriminatory capacity of the ORT in the prediction of an excessive response.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

DISCUSSION

The results of the present IPD meta-analysis, with data from 32 individual studies, demonstrate that both AFC and AMH clearly add value to female age alone in the prediction of excessive response. Antimüllerian hormone and AFC in concert have high predictive accuracy, even without adding female age. The results also indicate that the performance of the ORTs may vary across patient subgroups, as determined by female age especially. At a higher female age FSH performs less well, whereas AFC performs better in older age groups. Because FSH performs the least well in excessive response prediction, this finding is not very relevant. For AFC the change in predictive accuracy with increasing age is more notable and results in an increased predictive accuracy, in terms of an increase in the AUC of approximately 0.26. However, this increase is only seen with big increments of female age (from 20 to 30 years or 30 to 40 years). With smaller increases in female age, such as between 31, 34, and 37 years (the 25th, 50th, and 75th percentiles of age and thus the most clinically relevant group) the increase in AUC is much smaller and less clinically relevant. In addition, the gain

in predictive accuracy is evenly spread over the entirety of the curve, thus limiting the margin of additive clinical value.

The results of this IPD meta-analysis are mostly in line with those from a previous, conventional systematic review and meta-analysis of ovarian reserve tests and excessive response (35) and another recent study in which AMH was able to accurately identify 79% of excessive responders (57). Our IPD approach allowed us to evaluate the added value of ORTs on top of female age and, moreover, allowed for the analysis of accuracy in subgroups of women defined by age, BMI, or duration of subfertility. Although ORT adds value to female age in predicting excessive response, age adds little to nothing to the accuracy of the prediction based on the ORTs. It does, however, does seem to influence the accuracy of some ORTs.

The results of this IPD meta-analysis also suggest that age influences the accuracy of AFC and basal FSH. Although ovarian reserve decreases with age, the AFC is believed to reflect the true level of the quantitative ovarian reserve directly, in contrast to basal FSH, which constitutes an indirect marker of follicle numbers. Indeed, in older women the prevalence of excessive response may become too low for any test to gain sufficient accuracy, and this may be especially true for FSH. For AFC the change in accuracy may be significant only from the statistical point of view, without actual implications for clinical practice, and without an obvious explanatory mechanism.

A challenge with the IPD approach is collecting sufficient data. For the present study databases of 60 of the eligible 125 manuscripts were obtained. We were unable to reach a number of authors, primarily because of inaccurate contact information or because authors did not reply to the e-mail addresses provided. Older data were often lost or in a format that could no longer be read. Studies to investigate the possibility of combining IPD data with aggregated data are ongoing (58). To compare included and excluded studies, we aimed to calculate Spearman correlation coefficients for the included and nonincluded studies. Unfortunately, of the nonincluded studies only one reported sensitivity and specificity values for AFC in the prediction of an excessive response. Therefore, Spearman correlation could not be calculated. However, for 27 of 32 studies a Spearman correlation was calculated from a previous IPD meta-analysis on poor response prediction, and this showed that there was no difference (14). Because there is no difference in poor response prediction, it is reasonable to assume that there is also no difference for excessive response prediction. Therefore, we believe that the current number of participants and amount of data allowed us to analyze a valid selection of all the available data. It would have been interesting to add PCOS as a candidate predictor in our uni- and multivariate analyses because women with PCOS have been found to be prone to establishing OHSS after IVF treatment (13). However, in the majority of studies, PCOS was one of the exclusion criteria, and from those studies that included and recorded PCOS a mere 131 women had PCOS.

Although the present IPD meta-analysis included studies up to the end of 2009, the results of more recent studies on the value of ORTs in predicting ovarian response are still in agreement with our findings of the present IPD meta-analysis. Two recent studies in an IVF setting (57, 59) and three studies performed in oocyte donors or breast cancer patients undergoing oocyte cryopreservation all show an AUC of approximately 0.80 for AMH in excessive response prediction (60–62).

Using original data of a number of studies comes with between-study heterogeneity. The incorporation of ovarian reserve tests and restrictions based on test results in everyday IVF practice has led to selection bias in some study populations. Heterogeneity found in the included studies pertained to differences in IVF indications, access to IVF resources, differing treatment protocols, variability in embryo laws, and discordant definitions of ongoing pregnancy. There is also a variation in hormone assays and AFC sizes measured, for which no international consensus exists to correct for these differences. Consequently, no cutoff values for these tests could be used or mentioned. The most valuable method of obtaining such cutoff values for clinical practice is through randomized controlled trials, which are underway at the moment (63). We have used random intercept logistic regression as well as the ROC regression model by Janes, Pepe, and colleagues (41, 42), in which pertinent heterogeneity between studies is accounted for.

The clinical value of excessive response prediction will depend on the consequences for clinical management. Several studies have looked at the effect of individualized treatment protocols. By providing women with personally tailor-made stimulation protocols (i.e., with a lower daily FSH dosage), it is attempted to keep the oocyte yield between 5 and 12 oocytes. At present the evidence is inconclusive upon the effectiveness of such personalized treatment regimens based on a priori prediction of ovarian response (50, 51). In the study of Popovic-Todorovic et al. (51) the use of an individualized protocol resulted in a larger number of normal responders but a similar number of excessive responders. In contrast, Olivennes et al. (64) demonstrated that lower individualized dosage protocols allow for a similar oocyte yield, implantation rate, and pregnancy compared with higher dosage protocols. A third study (65) showed no difference in the number of mature oocytes retrieved or in the occurrence of OHSS between patients that were randomly assigned to receive 225 IU or 300 IU of FSH. Last, it has been suggested that individualization of stimulation protocols dose on the basis of ovarian reserve tests is expected to be cost effective in IVF populations (66).

On the basis of the present study we cannot speculate about associations between FSH dosage and excessive response prevention. A significant association between daily FSH dosage and excessive response was found, with women with lower daily FSH dosages having higher chances of excessive response. This association reflects physician behavior, whereby lower daily FSH dosages are preemptively prescribed according to specific patient characteristics, ORT results, or any comorbidity in anticipation of an excessive response. This suggests a form of selection bias, whereby the accuracy of ORTs or patient characteristics in the prediction of an excessive response is actually higher than currently reported, because some excessive responses may have been prevented by prescribing lower daily FSH dosages. The high response despite

a low daily FSH dosage can be explained by the presence of a large number of follicles, with a sensitivity for FSH close to the FSH threshold (67). More prospectively collected evidence, in the form of large-scale randomized control trials, is needed to demonstrate whether an individualized treatment protocol based on ORTs and patient characteristics is a truly effective strategy in the prevention of an excessive response; a protocol for such a randomized control trial was recently published (63).

In conclusion, this IPD meta-analysis shows that AFC and AMH add predictive accuracy to age in the prediction of an excessive response. A model combining these ORTs provides good predictive accuracy, without the necessity of including female age. The performance of FSH and AFC, but not AMH, was influenced by female age but not by BMI or duration of subfertility. However, the performance across subgroups with small increments in female age seemed not to be sufficiently altered to be recognized as clinically relevant. The high predictive accuracy for both AMH and AFC or a combination of both urges the need for studies that examine the effect of ORT-based dose adaptations in which efficacy of treatment, costs, and response normalization is analyzed.

Acknowledgments: The following contributions were made to the preparation of this article. Data collection: A. Aflatoonian, R. A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A. B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, T. Freour, C. Gnoth, E. M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, C. B. Lambalk, M. McIlveen, L. Mohiyiddeen, L. T. Merce, S. Muttukrishna, L. G. Nardo, S. M. Nelson, H. Y. Ng, B. Popovic-Todorovic, J. M. J. Smeenk, C. Tomás, P. J. Q. Van der Linden, I. A. van Rooij, I. K. Vladimirov, and F. J. M. Broekmans. Analyses: S. L. Broer, M. Dólleman, J. van Disseldorp, K. A. Broeze, B. C. Opmeer, P. Bossuyt, and M.J.C. Eijkemans. Interpretation of the results and writing of the article: S. L. Broer, M. Dólleman, J. van Disseldorp, K. A. Broeze, B. C. Opmeer, P. Bossuyt, M. J. C. Eijkemans, B. W. Mol, and F. J. M. Broekmans. Revision of the manuscript: all authors.

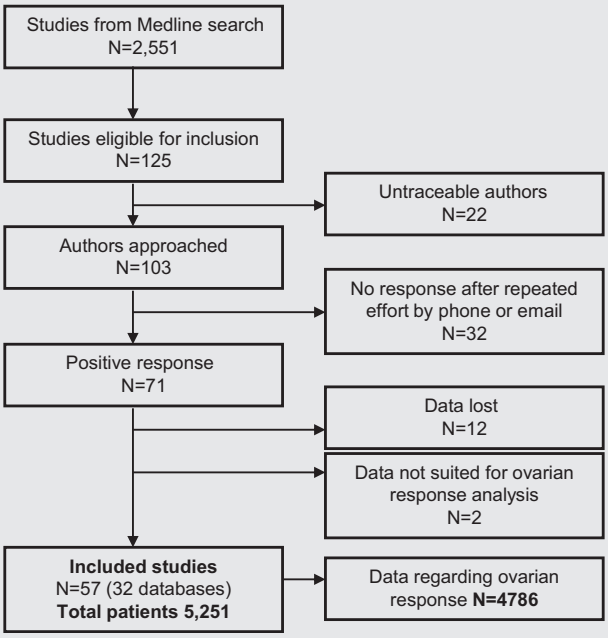
REFERENCES

1. Delvigne A, Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review. Hum Reprod Update 2002;8:559–77.
2. Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, et al. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil Steril 2000;73:552–7.
3. Baart EB, Martini E, van den Berg I, Macklon NS, Galjaard RJ, Fauser BC, et al. Preimplantation genetic screening reveals a high incidence of aneuploidy and mosaicism in embryos from young women undergoing IVF. Hum Reprod 2006;21:223–33.
4. 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;87:782–7.
5. Heijnen EM, Eijkemans MJ, De Klerk C, Polinder S, Beckers NG, Klinkert ER, et al. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. Lancet 2007;369:743–9.
6. Eldar-Geva T, Ben-Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, et al. Dynamic assays of inhibin B, anti-Mullerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum Reprod 2005;20:3178–83.
7. Verberg MF, Eijkemans MJ, Macklon NS, Heijnen EM, Baart EB, Hohmann FP, et al. The clinical significance of the retrieval of a low number of oocytes following mild ovarian stimulation for IVF: a meta-analysis. Hum Reprod Update 2009;15:5–12.
8. van der Gaast MH, Eijkemans MJ, van der Net JB, de Boer EJ, Burger CW, van Leeuwen FE, et al. Optimum number of oocytes for a successful first IVF treatment cycle. Reprod Biomed Online 2006;13:476–80.
9. 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.
10. 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.
11. Freour T, Mirallie S, Bach-Ngohou K, Denis M, Barriere P, Masson D. Measurement of serum anti-Mullerian hormone by Beckman Coulter ELISA and DSL ELISA: comparison and relevance in assisted reproduction technology (ART). Clin Chim Acta 2007;375:162–4.
12. Gnoth C, Schuring AN, Friol K, Tigges J, Mallmann P, Godehardt E. Relevance of anti-Mullerian hormone measurement in a routine IVF program. Hum Reprod 2008;23:1359–65.
13. Ho HY, Lee RK, Lin MH, Hwu YM. Estradiol level on day 9 as a predictor of risk for ovarian hyperresponse during controlled ovarian hyperstimulation. J Assist Reprod Genet 2003;20:222–6.
14. Broer SL, Mol B, Dolleman M, Fauser BC, Broekmans FJ. The role of anti-Mullerian hormone assessment in assisted reproductive technology outcome. Curr Opin Obstet Gynecol 2010;22:193–201.
15. Jayaprakasan K, Hilwah N, Kendall NR, Hopkisson JF, Campbell BK, Johnson IR, et al. Does 3D ultrasound offer any advantage in the pretreatment assessment of ovarian reserve and prediction of outcome after assisted reproduction treatment? Hum Reprod 2007;22:1932–41.
16. 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 2005;83:811–4.
17. van Rooij I, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH, et al. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002;17:3065–71.
18. 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;18:1422–7.
19. La Marca A, Giulini S, Tirelli A, Bertucci E, Marsella T, Xella S, et al. Anti-Mullerian hormone measurement on any day of the menstrual cycle strongly predicts ovarian response in assisted reproductive technology. Hum Reprod 2007;22:766–71.
20. 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;85:1541–3.
21. 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.
22. Smeenk JM, Stolwijk AM, Kremer JA, Braat DD. External validation of the templeton model for predicting success after IVF. Hum Reprod 2000;15:1065–8.
23. Riggs RM, Duran EH, Baker MW, Kimble TD, Hobeika E, Yin L, et al. 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;199:202–8.
24. Smeenk JM, Sweep FC, Zielhuis GA, Kremer JA, Thomas CM, Braat DD. Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. Fertil Steril 2007;87:223–6.

25. Nardo LG, Gelbaya TA, Wilkinson H, Roberts SA, Yates A, Pemberton P, et al. Circulating basal anti-Mullerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. Fertil Steril 2009;92:1586–93.
26. McIlveen M, Skull JD, Ledger WL. Evaluation of the utility of multiple endocrine and ultrasound measures of ovarian reserve in the prediction of cycle cancellation in a high-risk IVF population. Hum Reprod 2007;22:778–85.
27. 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.
28. Steinkampf MP, Hammond KR, Nichols JE, Slayden SH. Effect of obesity on recombinant follicle-stimulating hormone absorption: subcutaneous versus intramuscular administration. Fertil Steril 2003;80:99–102.
29. 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.
30. 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;91:705–14.
31. Broer SL, Eijkemans MJ, Scheffer GJ, van Rooij I, de Vet A, Themmen AP, et al. Anti-mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. J Clin Endocrinol Metab 2011;96:2532–9.
32. Broer SL. Assessment of current and future ovarian reserve status. Enschede, The Netherlands: Gildeprint Drukkerijen; 2011.
33. Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PM, Coppus SF, 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.
34. Broeze KA, Opmeer BC, Van Geloven N, Coppus SF, Collins JA, Den Hartog JE, et al. Are patient characteristics associated with the accuracy of hysterosalpingography in diagnosing tubal pathology? An individual patient data meta-analysis. Hum Reprod Update 2011;17:293–300.
35. Broer SL, Dolleman M, Opmeer BC, Fauser BC, Mol BW, Broekmans FJ. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2011;17:46–54.
36. Ebner T, Sommergruber M, Moser M, Shebl O, Schreier-Lechner E, Tews G. Basal level of anti-Mullerian hormone is associated with oocyte quality in stimulated cycles. Hum Reprod 2006;21:2022–6.
37. 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;26:319–25.
38. Ng EH, Yeung WS, Ho PC. The significance of antral follicle count in controlled ovarian stimulation and intrauterine insemination. J Assist Reprod Genet 2005;22:323–8.
39. Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. Hum Reprod 2011;26:1768–74.
40. Sterrenburg MD, Veltman-Verhulst SM, Eijkemans MJ, Hughes EG, Macklon NS, Broekmans FJ, et al. 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;17:184–96.
41. Janes H, Longton G, Pepe M. Accommodating covariates in ROC analysis. Stata J 2009;9:17–39.
42. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. Stata J 2009;9:1.
43. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
44. 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;91:53–7.
45. Caroppo E, Matteo M, Schonauer LM, Vizziello G, Pasquadibisceglie A, Vitti A, et al. Basal FSH concentration as a predictor of IVF outcome in older women undergoing stimulation with GnRH antagonist. Reprod Biomed Online 2006;13:815–20.
46. Muttukrishna S, Suharjono H, McGarrigle H, Sathanandan M. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? BJOG 2004;111:1248–53.
47. Muttukrishna S, McGarrigle H, Wakim R, Khadum I, Ranieri DM, Serhal P. Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? BJOG 2005;112:1384–90.
48. 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;22:2414–21.
49. 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;20:1647–54.
50. 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;18:781–7.
51. 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 2003;18:2275–82.
52. Tomas C, Nuojua-Huttunen S, Martikainen H. Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. Hum Reprod 1997;12:220–3.
53. van Swieten EC, van der Leeuw-Harmsen L, Badings EA, van der Linden PJ. Obesity and clomiphene challenge test as predictors of outcome of in vitro fertilization and intracytoplasmic sperm injection. Gynecol Obstet Invest 2005;59:220–4.
54. Vladimirov IK, Tacheva DM, Kalinov KB, Ivanova AV, Blagoeva VD. Prognostic value of some ovarian reserve tests in poor responders. Arch Gynecol Obstet 2005;272:74–9.
55. Vladimirov I, Tacheva D, Blagoeva V. [Prognostic value of some hormonal and ultrasound ovarian reserve tests]. Akush Ginekol (Sofiia) 2003;42:14–20.
56. 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.
57. Anckaert E, Smitz J, Schiettecatte J, Klein BM, Arce JC. The value of anti-Mullerian hormone measurement in the long GnRH agonist protocol: association with ovarian response and gonadotrophin-dose adjustments. Hum Reprod 2012;27:1829–39.
58. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 2008;27:6111–36.
59. Andersen AN, Witjes H, Gordon K, Mannaerts B. Predictive factors of ovarian response and clinical outcome after IVF/ICSI following a rFSH/GnRH antagonist protocol with or without oral contraceptive pre-treatment. Hum Reprod 2011;26:3413–23.
60. Lee S, Ozkavukcu S, Heytens E, Moy F, Alappat RM, Oktay K. Anti-Mullerian hormone and antral follicle count as predictors for embryo/oocyte cryopreservation cycle outcomes in breast cancer patients stimulated with letrozole and follicle stimulating hormone. J Assist Reprod Genet 2011;28:651–6.
61. Nakhuda GS, Douglas NC, Thornton MH, Guarnaccia MM, Lobo R, Sauer MV. Anti-Mullerian hormone testing is useful for individualization of stimulation protocols in oocyte donors. Reprod Biomed Online 2011;22(Suppl 1):S88–93.
62. Riggs R, Kimble T, Oehninger S, Bocca S, Zhao Y, Leader B, et al. Anti-Mullerian hormone serum levels predict response to controlled ovarian hyperstimulation but not embryo quality or pregnancy outcome in oocyte donation. Fertil Steril 2011;95:410–2.
63. van Tilborg TC, Eijkemans MJ, Laven JS, Koks CA, de Bruin JP, Scheffer GJ, et al. The OPTIMIST study: optimisation of cost effectiveness through individualised FSH stimulation dosages for IVF treatment. A randomised controlled trial. BMC Womens Health 2012;12:29.

64. Olivennes F, Howies CM, Borini A, Germond M, Trew G, Wikland M, et al. Individualizing FSH dose for assisted reproduction using a novel algorithm: the CONSORT study. Reprod Biomed Online 2011;22(Suppl 1):S73–82.
65. Jayaprakasan K, Hopkisson J, Campbell B, Johnson I, Thornton J, Raine-Fenning N. A randomised controlled trial of 300 versus 225 IU recombinant FSH for ovarian stimulation in predicted normal responders by antral follicle count. BJOG 2010;117:853–62.
66. Moolenaar LM, Broekmans FJ, van Disseldorp J, Fauser BC, Eijkemans MJ, Hompes PG, et al. Cost effectiveness of ovarian reserve testing in in vitro fertilization: a Markov decision-analytic model. Fertil Steril 2011;96:889–94.
67. Van der Meer M, Hompes PG, De Boer JA, Schats R, Schoemaker J. Cohort size rather than follicle-stimulating hormone threshold level determines ovarian sensitivity in polycystic ovary syndrome. J Clin Endocrinol Metab 1998;83:423–6.

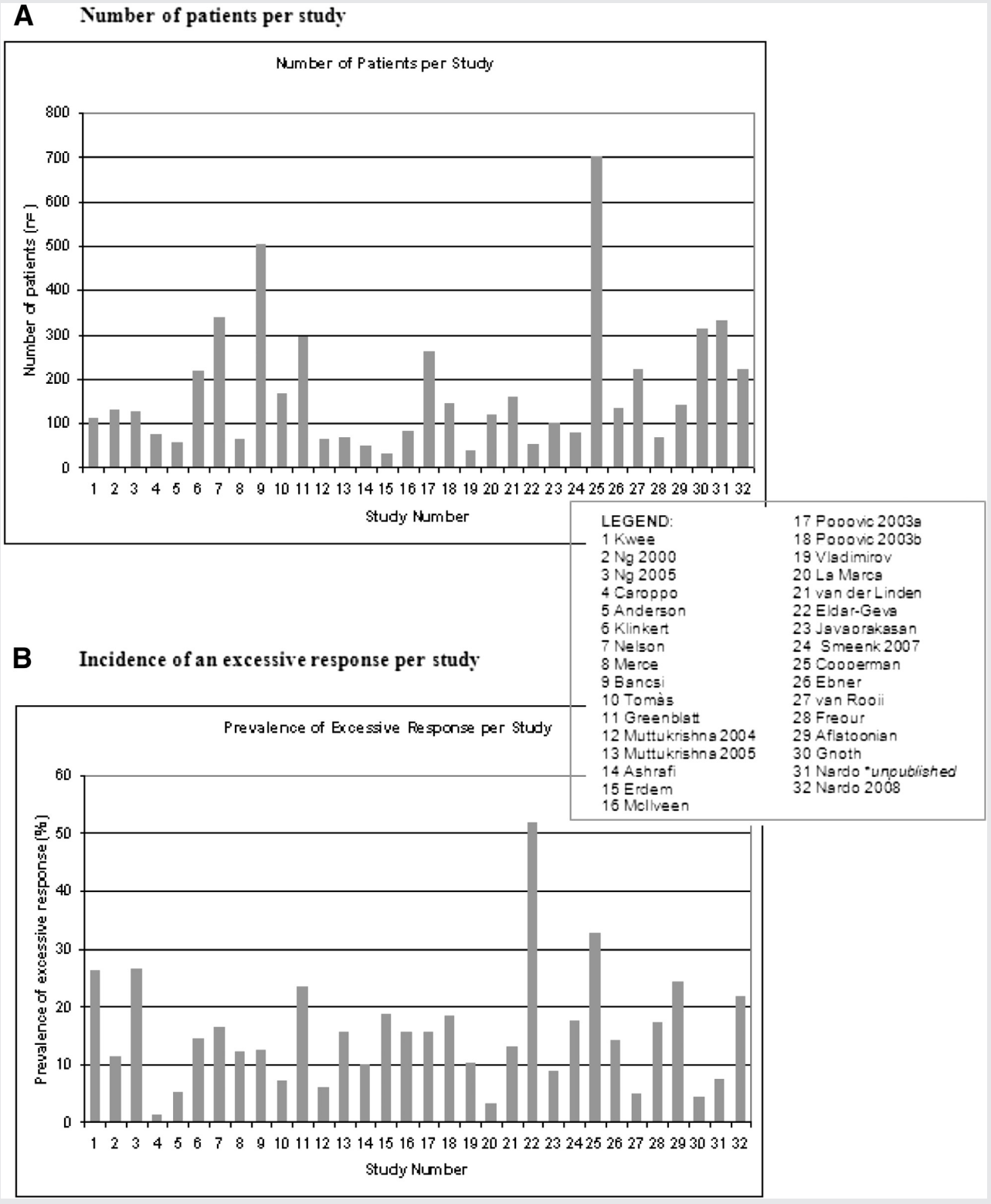
SUPPLEMENTAL FIGURE 1



Flowchart of included studies.

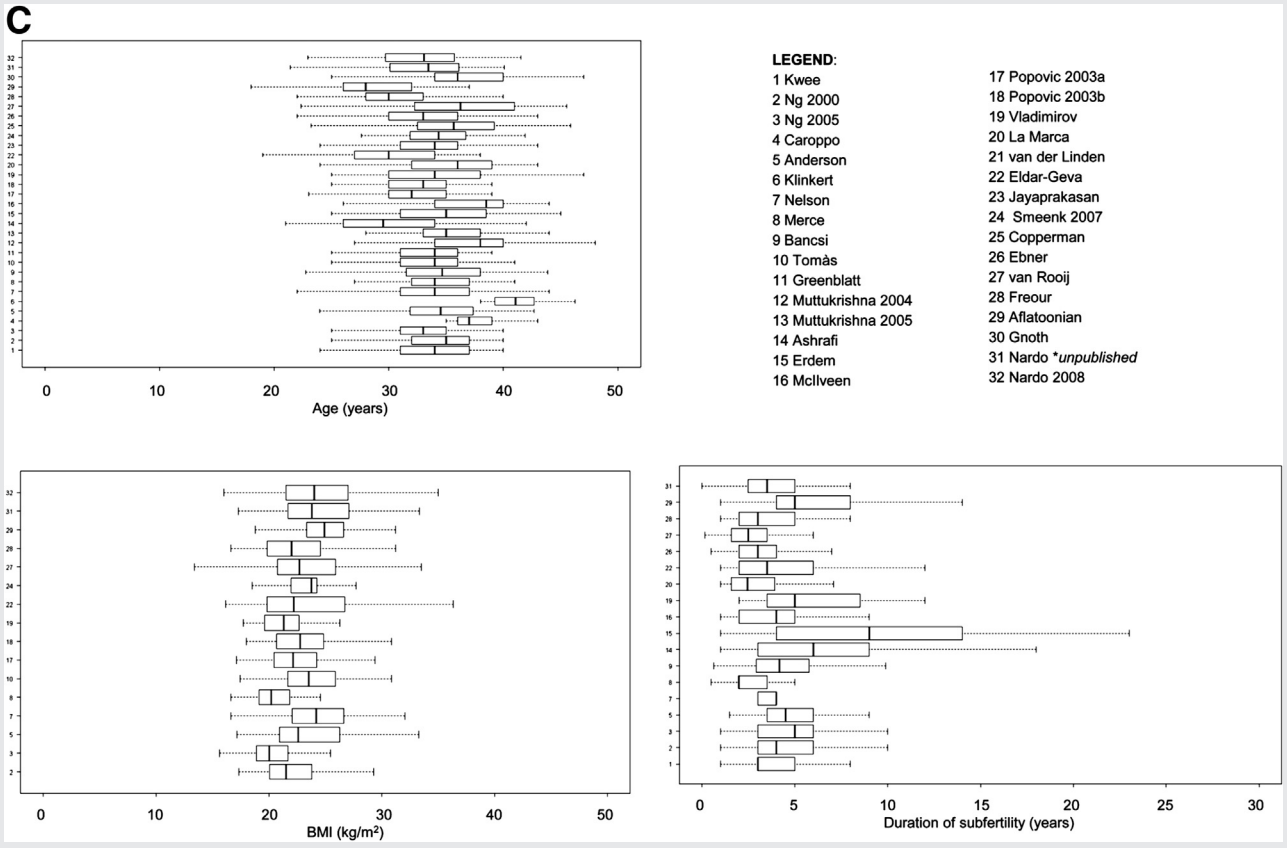
Broer. Excessive response prediction in IVF. Fertil Steril 2013.

SUPPLEMENTAL FIGURE 2



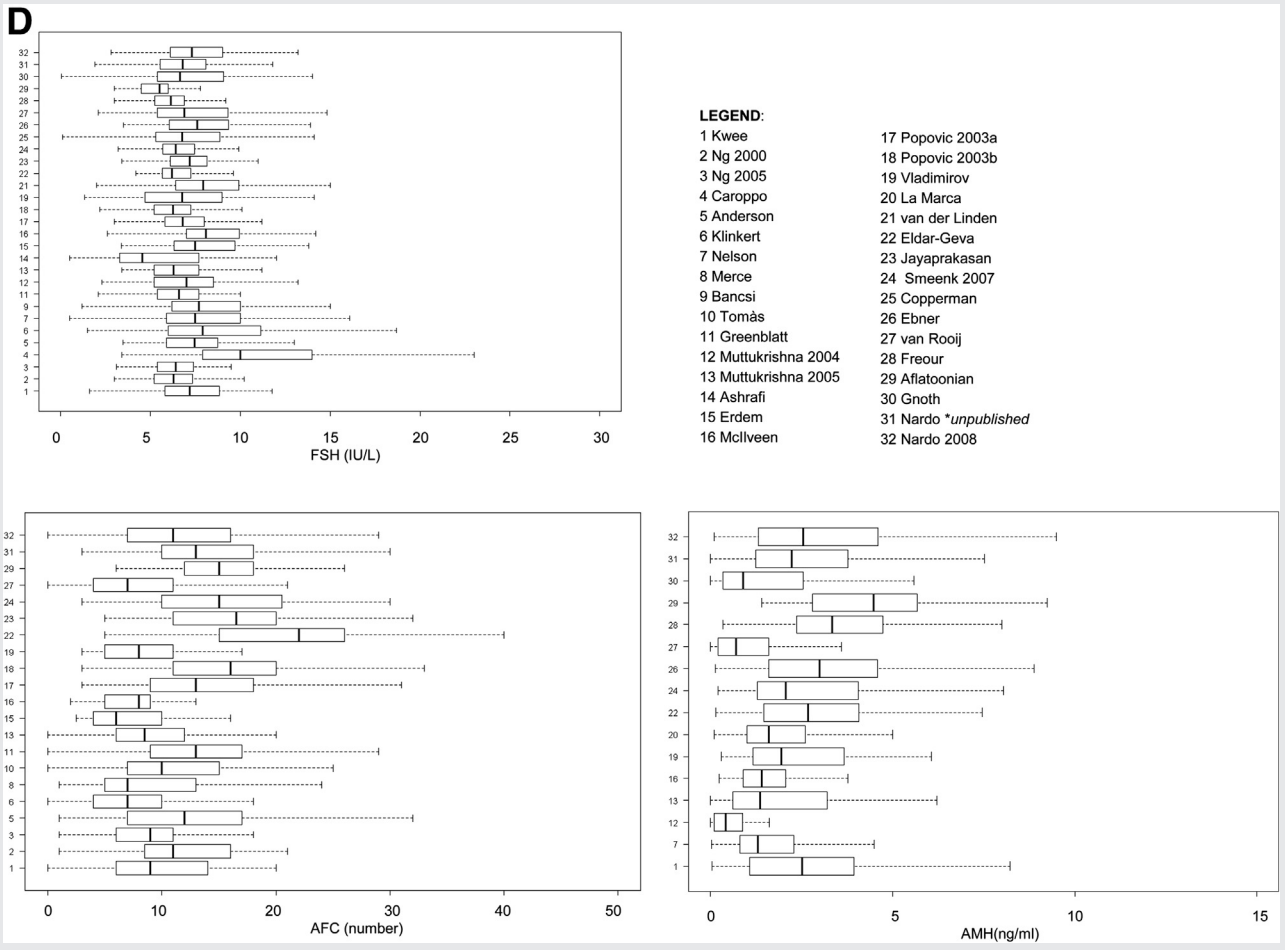
Broer. Excessive response prediction in IVF. Fertil Steril 2013.

SUPPLEMENTAL FIGURE 2 Continued



Broer. Excessive response prediction in IVF. Fertil Steril 2013.

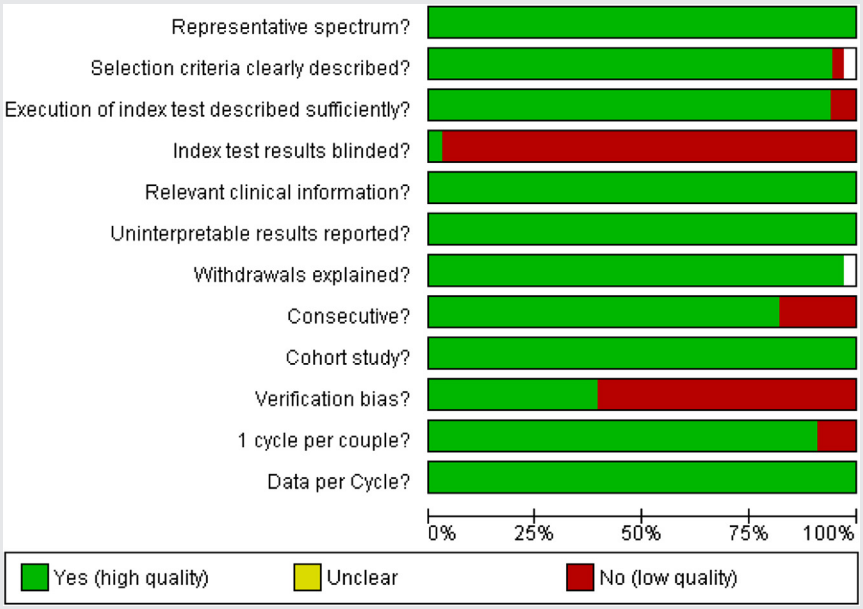
SUPPLEMENTAL FIGURE 2 Continued



Baseline characteristics of the included studies. (A) Number of patients per study. (B) Prevalence of an excessive response per study. (C) Mean, 5th percentile, and 95th percentile of the patient characteristics female age, BMI, and duration of subfertility for each individual study. (D) Mean, 5th percentile, and 95th percentile of ovarian reserve tests FSH, AFC, and AMH for each individual study.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

SUPPLEMENTAL FIGURE 3



Study characteristics according to QUADAS. 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, questions regarding 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, and some studies analyzed more cycles per couple, in which case only the first cycle was analyzed.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

SUPPLEMENTAL TABLE 1

AUCs of the included studies in the prediction of an excessive response.

	FSH			AFC			AMH		
Study	**AUC**	**N**	**Assay**	**AUC**	**N**	**Criteria (mm)**	**AUC**	**N**	**Assay**
Aflatoonian	0.60 (0.50–0.69)	143	12	0.96 (0.93–0.99)	143	2–6	0.94 (0.90–0.98)	143	DSL
Anderson	0.92 (0.99–1.00)	46	11	0.61 (0.67–0.85)	46	2–10	NA		
Ashrafi	0.59 (0.31–0.87)	50	NA	NA			NA		
Bancsi	0.61 (0.54–0.68)	505	6	NA			NA		
Caroppo	0.81 (0.72–0.90)	76	3	NA			NA		
Copperman	0.65 (0.60–0.69)	570	5	NA			NA		
Ebner	0.61 (0.46–0.75)	127	NA	NA			0.82 (0.74–0.90)	135	BC
Eldar-Geva	0.71 (0.57–0.85)	52	5	0.88 (0.75–1.00)	36	2–10	0.75 (0.62–0.88)	54	BC
Erdem	0.77 (0.57–0.97)	24	5	0.85 (0.70–1.00)	24	2–8	NA		
Freour	0.58 (0.41–0.73)	62	NA	NA			0.70 (0.55–0.86)	64	BC
Gnoth	0.64 (0.51–0.78)	122	NA	NA			0.87 (0.79–0.95)	134	DSL
Greenblatt	0.67 (0.59–0.74)	261	5	0.69 (0.61–0.77)	223	2–8	NA		
Jayaprakasan	0.74 (0.57–0.91)	100	NA	0.82 (0.70–0.95)	100	2–10	NA		
Klinkert	0.42 (0.30–0.55)	212	4	0.45 (0.33–0.57)	221	2–5	NA		
Kwee	0.79 (0.70–0.88)	109	1	0.87 (0.82–0.96)	109	2–10	0.84 (0.76–0.92)	105	DSL
La Marca	NA			NA			0.90 (0.76–1.00)	118	BC
McIlveen	No >15	71	8	No >15	71	2–10	No >15		BC
Merce	NA			0.62 (0.42–0.83)	65	2–5	NA		
Muttukrishna 1	0.81 (0.59–1.00)	66	7	NA			0.92 (0.83–1.00)	66	BC
Muttukrishna 2	0.67 (0.52–0.82)	68	7	0.84 (0.73–0.94)	68	NA	0.73 (0.56–0.91)	68	BC
Nardo 1	0.65 (0.53–0.77)	135	5	0.71 (0.59–0.83)	123	2–5	0.74 (0.64–0.83)	135	DSL
Nardo 2	0.68 (0.59–0.77)	145	13	0.71 (0.63–0.80)	145	2–5	0.79 (0.72–0.87)	145	DSL
Nelson	0.64 (0.58–0.71)	338	5	NA			0.88 (0.82–0.91)	319	DSL
Ng 1	0.70 (0.56–0.83)	131	2	0.80 (0.70–0.90)	131	NA	NA		
Ng 2	0.72 (0.56–0.83)	109	5	0.77 (0.68–0.85)	127	NA	NA		
Popovic 1	0.62 (0.54–0.71)	256	1	0.71 (0.63–0.80)	256	2–5	NA		
Popovic 2	0.62 (0.50–0.73)	143	1	0.76 (0.67–0.86)	143	2–5	NA		
Smeenk 1	0.54 (0.40–0.68)	80	10	0.66 (0.5300.79)	80	2–10	0.71 (0.57–0.84)	80	BC
Tomas	NA			0.82 (0.72–0.91)	160	2–5	NA		
Van Rooij	0.68 (0.58–0.79)	215	10	0.86 (0.79–0.93)	215	2–5	0.87 (0.77–0.97)	215	BC
Van der Linden	0.82 (0.72–0.92)	124	NA	NA			NA		
Vladimirov 2	0.67 (0.48–0.87)	39	9	0.74 (0.52–0.97)	39	2–10	0.80 (0.67–0.93)	39	BC

Note: FSH assays: 1 = Immunometric, Delfia; 2 = automated chemiluminescence, ACS180, Bayer; 3 = immunoradiometric, Immunotech; 4 = immunometric assay, Chiron Diagnostics; 5 = Immulite semiautomated, DPC; 6 = Enzymun-FSH test, Boehringer Mannheim; 7 = immunoradiometric assay, DPC; 8 = chemiluminescence detection, Adiva Centaur, Bayer; 9 = electrochemiluminescence immunoassay, Roche Elecsys; 10 = fluorescence immunoenzymometric, AxSYM, Abbott; 11 = double antibody assay, Organon; 12 = IDCS, Korbach; 13 = Roche E170 automated immunoassay. AMH assays: DSL = Diagnostic Systems Laboratories; BC = Beckman Coulter; NA = not available. Areas under the curve (AUC) from original studies for prediction models of ovarian reserve tests for the prediction of an excessive response.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.

SUPPLEMENTAL TABLE 2

Univariable and multivariable models of patient characteristics in the prediction of an excessive response.

	Three-test study group (n = 1,023)			Total study group (n = 4,786)		
Model	**OR**	**95% CI**	***P* value**	**OR**	**95% CI**	***P* value**
Univariable model						
Age	0.89	0.85–0.93	<.001	0.90	0.88–0.91	<.001
BMI	0.98	0.93–1.03	.405	1.00	0.97–1.03	.954
Duration	0.98	0.90–1.06	.555	0.97	0.92–1.01	.156
Multivariable model						
Age and BMI						
Age	0.91	0.87–0.95	<.001	0.9	0.87–0.93	<.001
BMI	0.99	0.93–1.04	.616	1.00	0.97–1.04	.976
Age and duration						
Age	0.90	0.85–0.94	<.001	0.89	0.86–0.91	<.001
Duration	1.01	0.93–1.10	.750	1.00	0.95–1.05	.956

Note: Duration = duration of subfertility.

Broer. Excessive response prediction in IVF. Fertil Steril 2013.