

# Prognostic models for high and low ovarian responses in controlled ovarian stimulation using a GnRH antagonist protocol

Frank J. Broekmans<sup>1,\*</sup>, Pierre J.M. Verweij<sup>2</sup>, Marinus J.C. Eijkemans<sup>3</sup>, Bernadette M.J.L. Mannaerts<sup>2</sup>, and Han Witjes<sup>2</sup>

<sup>1</sup>Division of Female and Baby, Department for Reproductive Medicine and Surgery, University Medical Center, Utrecht 3584 CX, The Netherlands <sup>2</sup>MSD, Oss 5342CC, The Netherlands <sup>3</sup>Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht 3508GA, The Netherlands

\*Correspondence address. E-mail: f.broekmans@umcutrecht.nl

Submitted on March 7, 2014; resubmitted on March 7, 2014; accepted on March 26, 2014

**STUDY QUESTION:** Can predictors of low and high ovarian responses be identified in patients undergoing controlled ovarian stimulation (COS) in a GnRH antagonist protocol?

**SUMMARY ANSWER:** Common prognostic factors for high and low ovarian responses were female age, antral follicle count (AFC) and basal serum FSH and LH.

**WHAT IS KNOWN ALREADY:** Predictors of ovarian response have been identified in GnRH agonist protocols. With the introduction of GnRH antagonists to prevent premature LH rises during COS, and the gradual shift in use of long GnRH agonist to short GnRH antagonist protocols, there is a need for data on the predictability of ovarian response in GnRH antagonist cycles.

**STUDY DESIGN, SIZE, DURATION:** A retrospective analysis of data from the Engage trial and validation with the Xpect trial. Prognostic models were constructed for high (> 18 oocytes retrieved) and low (< 6 oocytes retrieved) ovarian response. Model building was based on the recombinant FSH (rFSH) arm ( $n = 747$ ) of the Engage trial. Multivariable logistic regression models were constructed in a stepwise fashion ( $P < 0.15$  for entry). Validation based on calibration was performed in patients with equivalent treatment ( $n = 199$ ) in the Xpect trial.

**PARTICIPANTS/MATERIALS, SETTING, METHODS:** Infertile women with an indication for COS prior to IVF. The Engage and Xpect trials included patients of similar ethnic origins from North America and Europe who had regular menstrual cycles. The main causes of infertility were male factor, tubal factor and endometriosis.

**MAIN RESULTS AND THE ROLE OF CHANCE:** In the Engage trial, 18.3% of patients had a high and 12.7% had a low ovarian response. Age, AFC, serum FSH and serum LH at stimulation Day 1 were prognostic for both high and low ovarian responses. Higher AFC and LH were associated with an increased chance of high ovarian response. Older age and higher FSH correlated with an increased chance of low ovarian response. Region (North America/Europe) and BMI were prognostic for high ovarian response, and serum estradiol at stimulation Day 1 was associated with low ovarian response. The area under the receiver operating characteristic (ROC) curve (AUC) for the model for a high ovarian response was 0.82. Sensitivity and specificity were 0.82 and 0.73; positive and negative predictive values were 0.40 and 0.95, respectively. The AUC for the model for a low ovarian response was 0.80. Sensitivity and specificity were 0.77 and 0.73, respectively; positive and negative predictive values were 0.29 and 0.96, respectively. In Xpect, 19.1% of patients were high ovarian responders and 16.1% were low ovarian responders. The slope of the calibration line was 0.81 and 1.35 for high and low ovarian responses, respectively, both not statistically different from 1.0. In summary, common prognostic factors for high and low ovarian responses were female age, AFC and basal serum FSH and LH. Simple multivariable models are presented that are able to predict both a too low or too high ovarian response in patients treated with a GnRH antagonist protocol and daily rFSH.

**LIMITATIONS, REASONS FOR CAUTION:** Anti-Müllerian hormone was not included in the prediction modelling.

**WIDER IMPLICATIONS OF THE FINDINGS:** The findings will help with the identification of patients at risk of a too high or too low ovarian response and individualization of COS treatment.

**STUDY FUNDING/COMPETING INTERESTS:** Financial support for this study and the editorial work was provided by Merck, Sharp & Dohme Corp. (MSD), a subsidiary of Merck & Co. Inc., Whitehouse Station, NJ, USA. F.J.B. received a grant from CVZ to his institution; P.J.M.V. and H.W. are employees of MSD, and B.M.J.L.M. was an employee of MSD at the time of development of this manuscript.

**TRIAL REGISTRATION NUMBERS:** NCT 00696800 and NCT00778999.

**Key words:** predictive modelling / ovarian response / recombinant FSH / GnRH antagonist

## Introduction

In assisted reproduction treatment (ART) an optimal response to controlled ovarian stimulation (COS) is of crucial importance. Both too low an ovarian response and too high an ovarian response are associated with increased cancellation rates and lower pregnancy rates, and previous literature suggests an optimal range of oocytes below and above which outcomes are compromised (van der Gaast *et al.*, 2006; Sunkara *et al.*, 2011). A high ovarian response may also increase the risk of developing ovarian hyperstimulation syndrome (Papanikolaou *et al.*, 2006). For this reason it is clinically relevant to identify predictors of ovarian response that may enable clinicians to identify patients at risk of a too high or too low ovarian response and to individualize COS treatment for these patients (Fauser *et al.*, 2008). Moreover, such individualization could be more cost-effective as it could both increase the efficacy and reduce the costs of ART.

Many studies have been conducted in the field of ovarian response prediction during the last 10 years (Popovic-Todorovic *et al.*, 2003) and various predictors for low ovarian response have been proposed (Hendriks *et al.*, 2005; Verberg *et al.*, 2007). Broekmans *et al.* (2006) performed a systematic review of these tests and found that antral follicle count (AFC) and basal FSH had the best sensitivity and specificity for predicting low ovarian response, with the recent addition of anti-Müllerian hormone (AMH) as possibly the most reliable predictor (Broer *et al.*, 2009). More recently, predictors for a high ovarian response have also been identified, with AMH and AFC demonstrating similar sensitivity and specificity (Broer *et al.*, 2011). However, it should be noted that the majority of this research has been performed in the context of GnRH agonist protocols. The introduction of GnRH antagonists to prevent premature LH rises during COS and the gradual shift of current care from long GnRH agonist to short GnRH antagonist protocols (Kolibianakis *et al.*, 2006; Al-Inany *et al.*, 2011) have prompted the need for research on the predictability of ovarian response in GnRH antagonist cycles. A recent prospective study including patients with and without oral contraceptive pretreatment indicated that AMH and basal FSH are statistically significant predictors of both the number of oocytes retrieved and the occurrence of an excessive ovarian response, whereas AMH alone was the main predictor for low ovarian response (Nyboe Andersen *et al.*, 2011).

The aim of this paper is to identify prognostic factors for high and low ovarian responses in COS using the GnRH antagonist protocol. With the identified predictors, simple prognostic models for low and excessive response are constructed from which patient-specific probabilities for either outcome can be derived, as the basis for studies on FSH starting dose adjustment.

## Methods

The prognostic models for high and low ovarian responses presented in this paper were developed and validated in different data sets: model building was based on data from the Engage trial (Devroey *et al.*, 2009), whereas model validation was performed using data from the Xpect trial (Nyboe Andersen *et al.*, 2011). A high ovarian response was defined as the collection of > 18 oocytes at retrieval or cycle cancellation due to high ovarian response, according to trial protocol. A low ovarian response was defined as the retrieval of less than six oocytes or cycle cancellation due to low ovarian response, according to trial protocol.

### Data sets

Engage [NCT00696800] was a double-blind, randomized, non-inferiority trial assessing the ongoing pregnancy rates after one injection of 150 µg corifollitropin alfa during the first week of stimulation, compared with daily injections of 200 IU recombinant FSH (rFSH; Puregon Pen, N.V. Organon, The Netherlands) using a standard GnRH antagonist protocol (0.25 mg ganirelix, Orgalutran, N.V. Organon). The intention-to-treat population comprised 1506 subjects with a mean age of 31.5 years and body weight of 68.6 kg. Data from the rFSH arm (750 subjects) of this study were used to construct the models for predicting high and low ovarian responses. The data used in the current analyses reflect minor corrections to the previously published Engage trial data (Devroey *et al.*, 2009) (see corrigendum Devroey *et al.*, 2014).

Xpect [NCT00778999] was a multinational trial to identify prognostic factors for an ovarian response. Subjects were randomized to receive either OC pretreatment or no OC pretreatment prior to their COS cycle. A treatment regimen of 200 IU rFSH and 0.25 mg GnRH antagonist was applied during the COS cycle (i.e. the same as in the daily rFSH arm of the Engage study). The intention-to-treat population consisted of 408 subjects of similar age and body weight as in Engage (mean, 31.7 years and 64.8 kg, respectively). Data from the non-OC arm (199 subjects) were used to validate the models for high and low ovarian responses.

The two studies had similar inclusion and exclusion criteria which allowed only patients with regular menstrual cycles to be included and were conducted in the same time frame (2006–2007 for Engage and 2006–2008 for Xpect). Ethnicity was also similar in Engage (86.7% White, 3.6% Black, 2.8% Asian; 6.8% 'Other') and Xpect (91.5% White, 2.0% Black, 5.0% Asian; 1.5% 'Other'). Finally, both studies included subjects from Europe ( $n = 347$  and  $n = 101$  in the relevant arms of Engage and Xpect, respectively) as well as North America ( $n = 403$  and  $n = 98$  in Engage and Xpect, respectively). Validated immunoassays were performed at a central laboratory to measure serum levels of FSH, LH, inhibin B, estradiol (E<sub>2</sub>) and progesterone. Levels of FSH, LH, E<sub>2</sub> and progesterone were determined by time-resolved fluoroimmunoassay (AutoDelfia<sup>®</sup> immunofluorometric assay, PerkinElmer Life and Analytical Sciences, Brussels, Belgium) with a coefficient of variation of 10%. Detection limits were 0.25 IU/l, 0.6 IU/l, 49.9 pmol/l and 0.38 ng/ml

for FSH, LH, E<sub>2</sub> and progesterone, respectively. Serum inhibin B levels were determined by using a validated immunoassay by Diagnostic Systems Laboratories (DSL; Webster, TX, USA) with a coefficient of variation of 10% and a detection limit of 10.0 pg/ml. AMH was only measured in the Xpect trial. Since it was not measured in the Engage trial, AMH could not be considered for inclusion in the prognostic models in the present study.

## Model building

Model building was based on data from the rFSH arm of the Engage trial (Devroey et al., 2009). Since prognostic factors for a high ovarian response may be different from those for a low ovarian response, separate logistic regression models were constructed for these two end-points. Age was included in both models by default. Other candidate prognostic factors or covariates were as follows:

- Age at menarche (years).
- Average menstrual cycle length (days).
- Duration of infertility (years).
- Alcohol use (self-reported; yes/no).
- Smoking status (self-reported; yes/no).
- BMI at baseline (kg/m<sup>2</sup>).
- FSH at Day 1 of stimulation (IU/l).
- LH at Day 1 of stimulation (IU/l).
- E<sub>2</sub> at Day 1 of stimulation (pmol/l).
- Progesterone at Day 1 of stimulation (nmol/l).
- Inhibin B at Day 1 of stimulation (pg/ml).
- AFC at Day 1 of stimulation (number of follicles < 11 mm).
- Total ovarian volume (ml).
- Study region (North America versus Europe).
- Previous IVF/ICSI (yes/no).

For each candidate prognosticator, the association with a high or low ovarian response was assessed using the  $\chi^2$  test (i.e. the score test in a logistic regression model). After the inclusion of age, covariates were selected using forward selection ( $P < 0.15$  for entry). Backward elimination ( $P > 0.15$  for removal) confirmed the covariate selection for the final model. The number of subjects with missing values for the covariates selected in the final models was limited: 66 in Engage and 26 in Xpect. Missing data were mainly for hormones (54 and 26 subjects in Engage and Xpect, respectively). The fact of whether data were missing or not was not associated with a high or low ovarian response. All subjects were included in the final models with missing covariate values imputed using linear regression (with covariates for age and region), if applicable. No other imputation of missing data was performed, except for setting hormone levels below the lower limit of detection to 0.5 times than the lower limit (as is common practice). First-order interaction terms and quadratic terms were tested, but not found to be statistically significant.

For the final logistic regression model for a high or low ovarian response the receiver operating characteristic (ROC) curve was plotted and the area under the curve (AUC, or *c*-statistic) was calculated. The 'optimal' point on the ROC curve is the one that provides the best trade-off between sensitivity and specificity (i.e. the point that is closest in distance to the upper left-hand corner where sensitivity and specificity are equal to 1). Associated with this point is the 'optimal' probability cut-off that provides the best balance between false positives and false negatives for a high (or low) ovarian response. If the predicted probability for a given patient exceeded this optimal cut-off the patient was predicted to become a high (or low) ovarian responder, otherwise not. Sensitivity, specificity, positive predictive value and negative predictive value at the optimal cut-off were calculated.

These characteristics are data driven and presumably too optimistic. For this reason the calculated values were denoted as 'apparent' AUC, sensitivity, etc. Optimism-corrected values were calculated using leave-one-out cross-validation, i.e. the regression coefficients associated with the 'final

model' were re-estimated with each subject left out in turn. We then combined the 'leave-one-out' regression coefficient with the subject's covariate values in order to mimic the prediction of the outcome for each subject. Finally, a logistic regression model was fitted with the resulting 'leave-one-out' prognostic index (PI) as the only covariate in order to obtain the optimism-corrected AUC. Histograms displaying the distribution of the predicted probabilities were plotted separately for high or low ovarian responders and non-high (non-low) responders. Score charts (Hunault et al., 2004) were constructed for easier application of the two models.

## Model validation

A vital aspect of prediction is that a model derived from one data set can be transported to another. 'The idea of validating a prognostic model is generally taken to mean establishing that it works satisfactorily for patients other than those from whose data the model was derived' (Altman and Royston, 2000). External model validation was based on the non-OC arm of the Xpect study (Nyboe Andersen et al., 2011) and focused on two aspects: discrimination and calibration (Leushuis et al., 2009).

Discrimination is the ability of the model to distinguish between subjects with and without the event of interest, in this case between patients with a high or low ovarian response and patients without a high or low response. Discrimination was measured by the area under the ROC curve, the *c*-statistic. This statistic ranges from 0.5 (no discrimination) to 1 (perfect discrimination) and can be interpreted as the probability that for any discordant pair of subjects (i.e. one subject with the event and one without), the subject with the event has a higher predicted probability than the subject without the event (Harrell et al., 1996).

Calibration refers to correspondence between the predicted probabilities for a high or low ovarian response and the observed proportions. Calibration was assessed visually by comparing predicted probabilities and observed proportions after dividing patients in 10 groups based on their predicted probability and, more formally, by fitting a logistic regression model with a single covariate for the so-called PI, a linear combination of the subject's covariate values and the associated regression coefficients. Ideally, the regression coefficient of the PI is close to 1 and the intercept is close to 0. Usually the regression coefficient is < 1, indicating that the impact of the prognostic factors is less strong in new data: the well-known shrinkage phenomenon (Copas, 1983). An intercept different from 0 indicates that the overall event rate (in this case high and low ovarian responses, respectively) in the new data is different from the old data set.

All analyses were performed using SAS PC version 9.1. A  $P < 0.05$  was considered statistically significant.

## Results

Descriptive statistics for potential predictors are given in Tables I and II for the Engage and Xpect trials, respectively. Three patients in the Engage trial who discontinued their COS cycle due to an adverse event had a missing outcome and were excluded from the analysis, leaving 747 patients for analysis. A total of 137 patients had a high ovarian response and 95 patients had a low ovarian response, according to the definitions. In Xpect ( $n = 199$ ), there were 38 high responders and 32 low responders. The percentages of a high ovarian response in Engage and Xpect were similar (18.3 versus 19.1%), but the percentages of low responders were slightly different (12.7 versus 16.1%).

## Model building

### High ovarian response

In the Engage data the following factors had a strong ( $P < 0.001$ ) association with a high ovarian response (Table I): AFC at Day 1 of stimulation,

**Table 1** Descriptive statistics of potential predictors (covariates) for ovarian response in the rFSH arm of the Engage study—overall and by ovarian response category.

Covariate	Overall (n = 747)	Low (n = 95)	Normal (n = 515)	High (n = 137)	P-value*	
					High versus normal/low	Low versus normal/high
Age at baseline (years)						
Mean	31.5	32.8	31.7	30.2	<0.001	<0.001
SD	3.2	2.8	3.1	3.4		
Age at menarche (years)						
Mean	12.7	12.7	12.7	12.7	0.971	0.545
SD	1.3	1.4	1.3	1.3		
Average menstrual cycle length (days)						
Mean	28.5	28	28.4	28.8	0.020	0.016
SD	1.7	1.7	1.7	1.7		
Duration of infertility (years)						
Mean	3.2	3.3	3.2	3.2	0.901	0.731
SD	2.2	2.2	2.2	2.4		
Alcohol use (%)	42.3	38.9	44.3	37.2	0.148	0.563
Smoking (%)	8.9	7.4	9.1	8.8	0.987	0.584
BMI at baseline (kg/m <sup>2</sup> )						
Mean	24.8	25.1	24.7	25.2	0.199	0.292
SD	2.7	2.9	2.6	2.8		
Region (North America) (%)	53.7	54.7	48.9	70.8	<0.001	0.919
Race (White) (%)	86.7	88.4	87.4	83.2	0.579	0.266
Previous IVF/ICSI (%)	57.3	55.8	58.8	52.6	0.256	0.824
Cause of infertility**						
Male factor (%)	46.3	47.4	47	43.1	0.448	0.737
Tubal factor (%)	25.4	18.9	25.6	29.2	0.337	0.107
Endometriosis (%)	15.4	15.8	14	20.4	0.111	0.947
FSH at Day 1 of stimulation (IU/l) <sup>a</sup>						
Median	6.4	7.6	6.5	5.6	<0.001	<0.001
LH at Day 1 of stimulation (IU/l) <sup>a</sup>						
Median	4.4	4.1	4.5	4.6	0.043	0.608
E <sub>2</sub> at Day 1 of stimulation (pmol/l) <sup>a</sup>						
Median	119.3	123	119.3	114.9	0.384	0.042
Progesterone at Day 1 of stimulation (nmol/l) <sup>a</sup>						
Median	1.7	1.7	1.7	1.8	0.053	0.974
Inhibin B at Day 1 of stimulation (pg/ml) <sup>a</sup>						
Median	50.3	42.1	49.6	61.4	<0.001	0.003
AFC at Day 1 of stimulation (n)						
Mean	12.4	9.5	12.3	15.1	<0.001	<0.001
SD	4.5	9.5	12.3	15.1		
Total ovarian volume (ml) <sup>b</sup>						
Mean	13.2	11.9	12.7	15.8	<0.001	0.065
SD	7.1	11.9	12.7	15.8		
<sup>a</sup> n	693	90	478	125		
<sup>b</sup> n	627	77	440	120		

rFSH, recombinant FSH; E<sub>2</sub>, estradiol; AFC, antral follicle count.\*From the  $\chi^2$  score test in a logistic regression model.

\*\*Subjects could have more than one cause.

**Table II** Descriptive statistics of potential predictors for an ovarian response in the non-OC arm of the Xpect study (validation set)—overall and by ovarian response category.

Covariate	Overall (n = 199)	Low (n = 32)	Normal (n = 129)	High (n = 38)
Age at baseline (years)				
Mean	31.6	33.3	31.6	30.2
SD	4.1	3.3	4.3	3.9
Age at menarche (years)				
Mean	12.9	12.6	13.0	12.9
SD	1.5	1.6	1.5	1.5
Average menstrual cycle length (days)				
Mean	28.5	27.6	28.5	29.3
SD	1.8	1.4	1.8	1.7
Duration of infertility (years)				
Mean	3.7	3.8	3.7	3.4
SD	3.0	3.1	3.1	3.0
Alcohol use (%)	43.2	40.6	47.3	31.6
Smoking (%)	17.1	28.1	14.7	15.8
BMI at baseline (kg/m <sup>2</sup> )				
Mean	23.6	24.0	23.4	23.8
SD	3.4	4.3	3.3	2.9
Region (North America) (%)	49.2	37.5	47.3	65.8
Race (White) (%)	91.5	96.9	90.7	89.5
Previous IVF*	638	71.9	62.0	63.2
Cause of infertility				
Male factor (%)	55.3	56.3	57.4	47.4
Tubal factor (%)	19.6	15.6	20.2	21.1
Endometriosis (%)	9.0	9.4	10.1	5.3
FSH at Day I of stimulation (IU/l) <sup>a</sup>				
Median	6.7	8.1	6.7	5.5
LH at Day I of stimulation (IU/l) <sup>a</sup>				
Median	5.0	5.0	5.0	4.8
E <sub>2</sub> at Day I of stimulation (pmol/l) <sup>a</sup>				
Median	100.6	107.5	102.2	91.9
Progesterone at Day I of stimulation (nmol/l) <sup>a</sup>				
Median	1.6	1.7	1.6	1.5
Inhibin B at Day I of stimulation (pg/ml) <sup>a</sup>				
Median	47.9	25.3	49.7	57.2
AFC at Day I of stimulation (n)				
Mean	11.7	8.5	12.1	13.3
SD	5.9	3.3	5.8	6.7
Total ovarian volume (ml)				
Mean	12.0	9.4	12.0	14.1
SD	5.8	4.2	5.4	7.2
<sup>a</sup> n	173	25	114	34

OC, observed cases.

\*Subjects could have more than one cause.

FSH at Day I of stimulation, female age, total ovarian volume, study region and inhibin B. The multivariable logistic regression model (Table III) included female age, AFC Day I, FSH level Day I, LH level Day I, study region and BMI as independent predictors.

As shown in Table III, some factors that were not, or only marginally, statistically significant in the univariate analysis were still included in the multivariate model (e.g. BMI and LH). On the other hand, factors that were statistically significant when considered univariately (e.g. total

**Table III** Logistic regression model for a high ovarian response (>18 oocytes): stepwise-built logistic model, each row depicting the cumulative contribution of a variable to a model including all variables from previous rows.

Covariate	OR	95% CI	P-value	AUC <sup>a</sup>	AUC <sup>b</sup>
Age	0.89	0.83–0.95	0.0003	0.64	0.61
AFC	1.13	1.08–1.20	<0.0001	0.75	0.74
FSH	0.57	0.48–0.69	<0.0001	0.79	0.78
LH	1.26	1.11–1.46	0.0005	0.81	0.80
Region	2.24	1.44–3.49	0.0004	0.82	0.81
BMI	1.07	0.99–1.15	0.0890	0.82	0.81

Odds ratio (OR) for region is USA versus Europe. All other ORs are per unit increase. CI, confidence interval; AUC, area under the curve.

<sup>a</sup>Apparent.

<sup>b</sup>Optimism corrected.

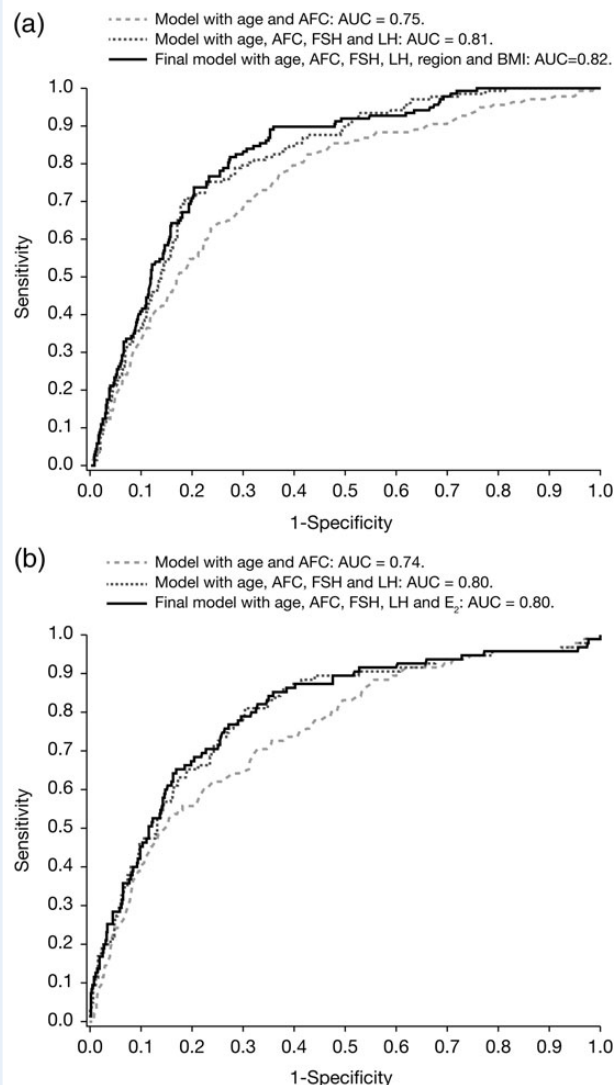
ovarian volume and inhibin B) were not included in the multivariate model. The prognostic impact of these factors was apparently captured by other factors already in the model. It appears that higher AFC, LH and BMI increased the chance of a high ovarian response, whereas higher FSH and older age decreased this risk. Also, a high ovarian response was more common in North America than in Europe.

More details of the model for a high ovarian response and application are given in the Supplementary data (see [Supplementary text 'Model formulas'](#) and [Supplementary Table S1](#)).

The apparent area under the ROC curve for a high ovarian response (Fig. 1a) was 0.82. The optimism-corrected AUC was only slightly lower (0.81). The optimal probability cut-off for the prediction of a high ovarian response was 17.9%. That is: if the model-based probability is higher than this value, a patient is classified as a 'predicted' high ovarian responder. The apparent sensitivity and specificity from this cut-off were 0.82 and 0.73, respectively. The apparent positive and negative predictive values were 0.40 and 0.95, respectively.

The discrimination achieved by models with fewer predictors was already close to that of the final model. A model with age, AFC, FSH and LH reached an AUC of 0.81. The ROC curve for this model was plotted in Fig. 1a. A model with only age and AFC, however, provided limited discriminatory capacity (AUC 0.75).

Histograms displaying the predicted probabilities for a high ovarian response based on the final model are given in the Supplementary data (see [Supplementary data, Fig. S1](#)). To assist in making model-based calculations in daily practice, a score chart was developed, together with a probability plot (Table IV, Fig. 2, for the model with four factors age, AFC, FSH and LH). The use of this chart is best illustrated by an example. Suppose we have a patient, aged 36 years with an AFC (2–10 mm) of 16, a basal FSH of 4.9 IU/l and a basal LH of 2.9 IU/l, using the score chart the total score for this patient can be calculated as 1 + 10 + 5 + 6 = 22. In the probability plot it can be seen that the predicted probability for this patient to become a high ovarian responder is ~13%. The 'optimal' probability cut-off for a high ovarian response (17.9%) approximately corresponds to a total score of 23. It should be noted that the score chart uses categorized covariates leading to some loss of information (apparent AUC 0.78 versus 0.81 for continuous covariates).



**Figure 1** (a). Receiver operating characteristic (ROC) curves for models for a high ovarian response (>18 oocytes) in controlled ovarian stimulation (COS) using a GnRH antagonist protocol. (b). ROC curves for models for a low ovarian response (<6 oocytes) in COS using a GnRH antagonist protocol.

Interpretation and application of the model would be further simplified if the continuous covariates age, AFC, FSH and LH were classified as 'high' or 'low', for example by using the median as a cut-off. However, it is well known that dichotomization of continuous covariates leads to loss of information. Indeed, the AUC of the simpler model drops to 0.77 (details not shown). Similarly, if we would simply count the number of risk factors present for each patient (0–6), the AUC of a model based on that count is only 0.74 (details not shown).

#### Low ovarian response

In the Engage data, FSH at Day 1 of stimulation, AFC at Day 1 of stimulation and age were strongly ( $P < 0.001$ ) related to low ovarian response (Table I). In the multivariable logistic regression model (Table V) female

**Table IV** Score chart for a high or low ovarian response.

Variable	High ovarian response		Low ovarian response		Score	
	Range <sup>a</sup>	Score	Range <sup>a</sup>	Score		
Age (years)	—	28	5	—	24	6
	29	31	4	25	28	7
	32	33	3	29	31	8
	34	35	2	32	33	9
	36	—	1	33	—	10
AFC	—	6	6	—	6	5
	7	8	7	7	7	4
	9	10	8	8	10	3
	11	13	9	11	13	2
	14	—	10	14	—	1
FSH (IU/l)	—	5.5	5	—	6	6
	5.5	6	4	6	6.5	7
	6	6.5	3	6.5	7.5	8
	6.5	7	2	7.5	8	9
	7	—	1	8	—	10
LH (IU/l)	—	4	6	—	4	5
	4	5	7	4	5	4
	5	6	8	5	6.5	3
	6	8	9	6.5	9	2
	8	—	10	9	—	1

<sup>a</sup>Lower limit excluded; upper limit included.

**Table V** Logistic regression model for a low ovarian response (<6 oocytes): stepwise-built logistic model, each row depicting the cumulative contribution of a variable to a model including all variables from previous rows.

Covariate	OR	95% CI	P-value	AUC <sup>a</sup>	AUC <sup>b</sup>
Age	1.08	1.00–1.18	0.0560	0.63	0.58
AFC	0.87	0.82–0.93	<0.0001	0.75	0.74
FSH	1.47	1.28–1.68	<0.0001	0.78	0.77
LH	0.81	0.69–0.95	0.0085	0.80	0.78
E <sub>2</sub>	1.01	1.00–1.01	0.0454	0.80	0.78

OR are per unit increase.

<sup>a</sup>Apparent.

<sup>b</sup>Optimism corrected.

More details of the model for a low ovarian response and application are given in the Supplementary data (see [Supplementary text 'Model formulas'](#) and [Supplementary data, Table SII](#)).

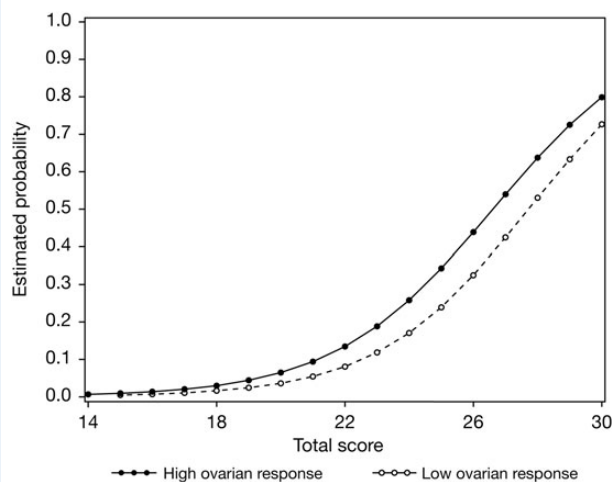
The apparent AUC of the ROC curve for the complete model (Fig. 1b) was 0.80. The optimal probability cut-off for the prediction of a low ovarian response was 12.8% (i.e. a patient is classified as a predicted low ovarian responder if the model-based probability is above this value). The apparent sensitivity and specificity for this cut-off level were 0.77 and 0.73, respectively. The apparent positive and negative predictive values were 0.29 and 0.96, respectively. Again, it appeared that the discrimination achieved by a simpler model was close to that of the complete final model (Table V). A model with age, AFC, FSH and LH already achieved an AUC of 0.80. The ROC curve for this model is plotted in Fig. 1b.

Histograms with the predicted probabilities for a low ovarian response are given in the Supplementary data (see [Supplementary Fig. S2](#)). A score chart was also provided for a low ovarian response (Table IV, again for the model with the four factors age, AFC, FSH and LH). It should be noted that for the same variable, the categorizations and scores are different from the score chart for high response. Continuing the example of the 36-year-old patient, the total score for this patient can be calculated as 10 + 1 + 6 + 5 = 22. In the probability plot (Fig. 2) it can be seen that the predicted probability for this patient to become a low ovarian responder is <10%. The 'optimal' probability cut-off for a low ovarian response (12.8%) approximately corresponds to a total score of 23. Note, again, that some information is lost due to categorization of covariates in the score chart (apparent AUC 0.78 versus 0.80).

Again, the interpretation of the model could be further simplified by classifying the covariates as 'high' or 'low' based on their median values. However, the AUC of the simpler model would then drop to 0.73 (details not shown). Similarly, the AUC of a model based on the number of risk factors present (0–5) would become 0.71 (details not shown).

## Model validation

A calibration plot for a high ovarian response (see [Supplementary Fig. S3](#)) demonstrated that there was reasonable agreement between the observed percentages in the Xpect data and the predicted probabilities

**Figure 2** Probability plot for a high or low ovarian response in COS using a GnRH antagonist protocol.

age, AFC Day I, basal FSH level, basal LH level and E<sub>2</sub> on Day I were included as independent predictors.

Four prognostic factors identified for a low ovarian response were also identified for a high ovarian response. As expected, the direction of the effects was reversed: higher FSH and older age increased the chance of a low ovarian response, whereas higher AFC and LH decreased this risk.

based on the model derived from the Engage trial. A logistic regression model for a high ovarian response in the Xpect data with the PI as the only covariate resulted in a regression coefficient of 0.81, smaller than unity but not statistically significantly so ( $P = 0.26$ ). The intercept was virtually zero ( $P = 0.98$ ), indicating that, corrected for the PI, the percentage of high responders was well predicted. The associated AUC was 0.78, smaller than the apparent AUC (0.82).

The calibration plot for a low ovarian response (see [Supplementary Fig. S4](#)) showed again agreement between predicted and observed percentages, except for one outlier. Surprisingly, the regression coefficient of the PI for a low ovarian response was greater than 1 (1.35), although the difference from unity was not statistically significant ( $P = 0.18$ ). The associated AUC was 0.84, in fact, greater than the apparent AUC of 0.80, suggesting an increased ability to distinguish patients, something that is not observed very often in prognostic modelling. The intercept was 0.77 ( $P = 0.090$ ) suggesting that, when corrected for the PI, the percentage of low responders in Xpect was underestimated. Apparently, the model could not fully explain the difference in low responder rates between Engage (12.7%) and Xpect (16.1%).

### Model building and validation using a model for a high ovarian response based on the number of follicles

Model building and validation using a definition of a high ovarian response as  $> 18$  follicles  $\geq 11$  mm diameter on the day of hCG administration are given in the Supplementary data (see [Supplementary text 'Alternative model for a high ovarian response based on the number of follicles'](#), [Supplementary data, Table SIII and Figs S5 and S6](#)).

## Discussion

The present study confirms the ability of prior prediction of high and low responders to COS using a GnRH antagonist for LH rise prevention. The common prognostic factors for high and low ovarian responses were female age, AFC and basal serum FSH and LH. In conjunction, these factors provide sufficiently accurate response prediction models for studies on individualized tailoring of the FSH stimulation dosage.

The importance of AFC and basal FSH, as well as female age, is in line with data from long GnRH agonist protocols ([Broekmans et al., 2006](#); [Fauser et al., 2008](#); [Broer et al., 2009](#)). Although AFC and basal FSH may both relate to the quantity of FSH-sensitive follicles, their independent contribution to at least the prediction of low response has been demonstrated in several studies ([Verhagen et al., 2008](#)). The estimate of overall sensitivity and specificity of published prediction models for a low ovarian response, based on the summary ROC curve in a published meta-analysis ([Verhagen et al., 2008](#)), clearly matched the findings for the currently presented model. For exaggerated response prediction, formal multifactor prediction models have not been published, as most of the attention has focused on single-test predictors, such as AMH and AFC ([Broer et al., 2011](#)).

The association between LH and ovarian hypo- and hyper-response has not been identified previously. A limited number of studies have included LH levels in an LH/FSH ratio, with the purpose of assessing its value for outcome prediction ([Mukherjee et al., 1996](#); [Shrim et al., 2006](#)). However, a formal meta-analysis of these studies is lacking, and its value seems limited. The association between elevated LH levels

and polycystic ovary syndrome may explain the current findings, although a more linear relation with the number of antral follicles is clearly absent for this factor.

The inclusion of study region in the model for a high ovarian response improves predictions, but lacks any biological rationale, other than a possible imbalance in predictive factors between European and North American populations. Therefore, we investigated whether the region effect could be explained by other factors. It appeared that there were differences between regions, but only for covariates that were not included in the model: smoking status (Europe versus North America: 13.6 versus 4.8%), serum progesterone at Day 1 of stimulation (median 1.6 versus 1.8 nmol/l) and total ovarian volume (median 9.5 versus 13.7 ml). Forced inclusion of these factors in the model did not eliminate the effect of study region. The only remaining explanation is that study region captures differences in variables that have not been specifically recorded, for example the oocyte retrieval procedure.

The fact that the present findings and those of a previous report ([Nyboe Andersen et al., 2011](#)) clearly confirms the predictability of ovarian response categories in antagonist co-treatment cycles is an important finding. In view of the differences in the way the ovaries are exposed to exogenous FSH, the possibility was expressed that submaximal stimulation could undermine the predictability by factors such as AMH and AFC. Assuming that these factors would correctly indicate the number of FSH-sensitive follicles, increased variation in the proportion of follicles that will indeed grow and deliver an oocyte in antagonist cycles could create a possible source for inaccuracy. Apparently, the proportional relation between cohort size at initiation of stimulation and the oocyte yield at the end of the track is not different when agonist and antagonist cycles are compared, though a systematic difference in oocyte yield has been firmly demonstrated for these two treatment approaches ([Al-Inany et al., 2011](#)).

No uniform definitions were available for excessive and a low ovarian response at the time of writing of this paper. We have used  $> 18$  and  $< 6$  oocytes for high and low ovarian responses, respectively ([Ferraretti et al., 2011](#)). Alternative definitions for high ovarian ( $> 15$  rather than  $> 18$  oocytes) and low ovarian responses ( $< 5$  rather than  $< 6$  oocytes) were explored, but the same variables were selected with similar regression coefficients (results not shown). The best operative definition for either response type ultimately depends on the way a diagnostic category (for example 'low responder') will lead to a certain change in management. Current understanding points towards the range of 6–14 oocytes as the range of optimal response associated with the highest probability of a live birth ([Sunkara et al., 2011](#)). Certainly, the optimal limits may further be affected by the risk of complications, such as ovarian hyperstimulation syndrome, and the likelihood that, in cases with a predicted response outside of this range, adjusted management can alter the outcome to a response in the normal range. Expectations here may be more optimistic regarding prevention of an excessive response than for a low response ([Klinkert et al., 2005](#); [Lekamge et al., 2008](#); [Olivennes, 2010](#); [Jayaprakasan et al., 2012](#); [Nelson et al., 2012](#)).

The strength of the prediction models presented here is that both were validated in an independent study, showing good discrimination and calibration in a cohort of comparable patients. The prediction model included both FSH and LH, which were both consistently measured by a central laboratory using the same immunoassays. Due to the well-known differences between commercial gonadotrophin immunoassays, the external value of the model may become slightly different if



other commercial FSH and LH assays are applied. A weakness is the absence in the models of AMH, a factor that had a high prognostic value in agonist cycles (Broer et al., 2011). When modelling high and low response based on the Xpect study, where AMH was collected, this parameter turned out to be predictive for both high and low ovarian responses, replacing AFC in the models (Nyboe Andersen et al., 2011).

Although AMH has appeared to be a solid biomarker of ovarian reserve with a considerable degree of intra- and inter-cycle consistency (Hehenkamp et al., 2006; van Disseldorp et al., 2010), the AMH assay suffers from a certain degree of variability that may hamper reliable predictions of ovarian response (Rustamov et al., 2012).

One of the sources of this variation is the between-sample variation during one or subsequent menstrual cycles. This variation has appeared to be quite substantial, specifically in younger women (Overbeek et al., 2012; Rustamov et al., 2012) and is believed to represent biological fluctuation parallel to fluctuation in antral follicle numbers (van Disseldorp et al., 2010). Moreover, nomograms or prognostic models should be based on studies where the samples have been measured by the same AMH immunoassay to ensure accurate predictions (Nelson and La Marca, 2011).

Based on the present findings and studies in agonist cycles, AMH and AFC may serve as highly overlapping predictors, with currently no definite conclusion as to the factor with the highest performance (Broer et al., 2011).

The lack of AMH as a factor in the model may not be permanent. Prognostic models may be updated when new predictors or tests become available and techniques for quick updating (as opposed to extensive model revisions) exist (Steyerberg et al., 2004). Another large trial in patients undergoing COS using a GnRH antagonist protocol has been completed recently [Pursue (NCT01144416)]. Since this trial is similar to Engage in design and sample size and includes AMH assessments, an update of the presented models may be indicated in due course.

## Implications for practice

The usefulness of ovarian response prediction for clinical practice will depend on two issues. First, the accuracy of the response class prediction needs to limit the number of false predictions. For the models presented here, ~75% of real low or high responders can be identified; however, at the same time, a positive test will, in some 15% of cases, wrongly suggest that the patient is producing too few or too many oocytes. It is crucial to consider that cases with a normal test will receive standard treatment, while cases with abnormal tests will be managed differently, for example, by dosage increase or dosage reduction. Secondly, dose reduction may create low response in falsely predicted high responders, while dose increase in falsely predicted low responders may create excessive responses. To what extent this will affect the overall efficacy of prior response predicting and subsequent adjustments in the stimulation regimen must be assessed from well-powered randomized trials. In such trials, both the efficacy of adjusted treatment in normalizing response and the effect of inaccuracies of prediction will be combined. Relevant outcome measures, such as overall programme performance, cancellation rates and costs, will in concert help to determine the true value of treatment individualization based on response prediction. Published scenario studies to date were non-randomized or not well

controlled (Olivennes, 2010; Nardo et al., 2011; Nelson et al., 2012). Currently executed studies will help to define the desired added value of tailored stimulation protocols (van Tilborg et al., 2012).

## Summary

Prognostic models to predict poor or excessive ovarian response in antagonist co-medicated ovarian hyperstimulation treatment for IVF appear to be as accurate as in agonist controlled cycles. This finding opens avenues for trials on individualized treatment protocols.

## Supplementary data

Supplementary data are available at <http://humrep.oxfordjournals.org/>.

## Authors' roles

F.J.B., P.J.M.V., M.J.C.E., B.M.J.L.M. and H.W. took part in the analysis and interpretation of data, writing the manuscript and in the final approval of the version to be published.

## Funding

Financial support for this study was provided by Merck, Sharp & Dohme Corp., a subsidiary of Merck & Co. Inc., Whitehouse Station, NJ, USA. Medical writing and editorial assistance was provided by P. Milner, PhD, of PAREXEL, UK. This assistance was funded by Merck, Sharp & Dohme Corp., a subsidiary of Merck & Co. Inc., Whitehouse Station, NJ, USA. Funding to pay the Open Access publication charges for this article was provided by Merck & Co., Inc., Whitehouse Station, NJ.

## Conflict of interest

F.J.B.: grant to his institution from CVZ. M.J.C.E.: none. P.J.M.V. and H.W. are employees of Merck, Sharp & Dohme Corp. (MSD) and B.M.J.L.M. was an employee of MSD.

## References

- Al-Inany HG, Youssef MA, Aboulghar M, Broekmans F, Sterrenburg M, Smit J, Abou-Setta AM. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. *Cochrane Database Syst Rev* 2011; CD001750.
- Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19**:453–473.
- 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, 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.
- Copas J. Regression, prediction and shrinkage. *J R Stat Soc Series B Stat Methodol* 1983; **45**:311–354.

- Devroey P, Boostanfar R, Koper NP, Mannaerts BM, Ijzerman-Boon PC, Fauser BC. A double-blind, non-inferiority RCT comparing corifollitropin alfa and recombinant FSH during the first seven days of ovarian stimulation using a GnRH antagonist protocol. *Hum Reprod* 2009;**24**:3063–3072.
- Devroey P, Boostanfar R, Koper NP, Mannaerts BMJL, Verweij PJM, Stegmann BJ, Ijzerman-Boon PC, Fauser BCJM, on behalf of the ENGAGE Investigators. Corrigendum: a double-blind, non-inferiority RCT comparing corifollitropin alfa and recombinant FSH during the first seven days of ovarian stimulation using a GnRH antagonist protocol. *Hum Reprod* 2014; published online 6 Mar 2014: doi 10.1093/humrep/deu030.
- 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.
- Ferraretti AP, La MA, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. *Hum Reprod* 2011;**26**:1616–1624.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–387.
- Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, te Velde ER, Broekmans FJ. Anti-Mullerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. *J Clin Endocrinol Metab* 2006;**91**:4057–4063.
- Hendriks DJ, Mol BW, Bancsi LF, te Velde ER, Broekmans FJ. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. *Fertil Steril* 2005;**83**:291–301.
- Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL, te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019–2026.
- Jayaprakasan K, Chan Y, Islam R, Haoula Z, Hopkisson J, Coomarasamy A, Raine-Fenning N. Prediction of in vitro fertilization outcome at different antral follicle count thresholds in a prospective cohort of 1,012 women. *Fertil Steril* 2012;**98**:657–663.
- 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* 2005;**20**:611–615.
- Kolibianakis EM, Collins J, Tarlatzis B, Papanikolaou E, Devroey P. Are endogenous LH levels during ovarian stimulation for IVF using GnRH analogues associated with the probability of ongoing pregnancy? A systematic review. *Hum Reprod Update* 2006;**12**:3–12.
- 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.
- Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol BW, Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**15**:537–552.
- Mukherjee T, Copperman AB, Lapinski R, Sandler B, Bustillo M, Grunfeld L. An elevated day three follicle-stimulating hormone:luteinizing hormone ratio (FSH:LH) in the presence of a normal day 3 FSH predicts a poor response to controlled ovarian hyperstimulation. *Fertil Steril* 1996;**65**:588–593.
- Nardo LG, Fleming R, Howles CM, Bosch E, Hamamah S, Ubaldi FM, Hugues JN, Balen AH, Nelson SM. Conventional ovarian stimulation no longer exists: welcome to the age of individualized ovarian stimulation. *Reprod Biomed Online* 2011;**23**:141–148.
- Nelson SM, La Marca A. The journey from the old to the new AMH assay: how to avoid getting lost in the values. *Reprod Biomed Online* 2011;**23**:411–420.
- Nelson SM, Anderson RA, Broekmans FJ, Raine-Fenning N, Fleming R, La MA. Anti-Mullerian hormone: clairvoyance or crystal clear? *Hum Reprod* 2012;**27**:631–636.
- Nyboe Andersen A, 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**:3413–3423.
- Olivennes F. Ovarian hyperstimulation syndrome prevention strategies: individualizing gonadotropin dose. *Semin Reprod Med* 2010;**28**:463–467.
- Overbeek A, Broekmans FJ, Hehenkamp WJ, Wijdeveld ME, van Disseldorp J, van Dulmen-den BE, Lambalk CB. Intra-cycle fluctuations of anti-Mullerian hormone in normal women with a regular cycle: a re-analysis. *Reprod Biomed Online* 2012;**24**:664–669.
- Papanikolaou EG, Pozzobon C, Kolibianakis EM, Camus M, Tournaye H, Fatemi HM, Van SA, Devroey P. Incidence and prediction of ovarian hyperstimulation syndrome in women undergoing gonadotropin-releasing hormone antagonist in vitro fertilization cycles. *Fertil Steril* 2006;**85**:112–120.
- Popovic-Todorovic B, Loft A, Lindhard A, Bangsboll S, Andersson AM, Andersen AN. A prospective study of predictive factors of ovarian response in 'standard' IVF/ICSI patients treated with recombinant FSH. A suggestion for a recombinant FSH dosage normogram. *Hum Reprod* 2003;**18**:781–787.
- Rustamov O, Smith A, Roberts SA, Yates AP, Fitzgerald C, Krishnan M, Nardo LG, Pemberton PW. Anti-Mullerian hormone: poor assay reproducibility in a large cohort of subjects suggests sample instability. *Hum Reprod* 2012;**27**:3085–3091.
- Shrim A, Elizur SE, Seidman DS, Rabinovici J, Wiser A, Dor J. Elevated day 3 FSH/LH ratio due to low LH concentrations predicts reduced ovarian response. *Reprod Biomed Online* 2006;**12**:418–422.
- Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;**23**:2567–2586.
- 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**:1768–1774.
- van der Gaast MH, Eijkemans MJ, van der Net JB, de Boer EJ, Burger CW, van Leeuwen FE, Fauser BC, Macklon NS. Optimum number of oocytes for a successful first IVF treatment cycle. *Reprod Biomed Online* 2006;**13**:476–480.
- van Disseldorp J, Lambalk CB, Kwee J, Looman CW, Eijkemans MJ, Fauser BC, Broekmans FJ. Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. *Hum Reprod* 2010;**25**:221–227.
- van Tilborg TC, Eijkemans MJ, Laven JS, Koks CA, de Bruin JP, Scheffer GJ, van Golde RJ, Fleischer K, Hoek A, Nap AW *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.
- Verberg MF, Eijkemans MJ, Macklon NS, Heijnen EM, Fauser BC, Broekmans FJ. Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF. *Hum Reprod* 2007;**22**:1919–1924.
- 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.