Human Reproduction, Vol.26, No.2 pp. 414–422, 2011
Advanced Access publication on December 8, 2010 doi:10.1093/humrep/deq339

human reproduction

ORIGINAL ARTICLE *Reproductive endocrinology*

Multi-marker assessment of ovarian reserve predicts oocyte yield after ovulation induction

Majedah Al-Azemi [1,*], Stephen R. Killick [2], Sheila Duffy [1], Clare Pye [1], Bassem Refaat [1], Natalie Hill [1], and William Ledger [1]

[1]Academic Unit of Reproductive and Developmental Medicine, Level 4, The Jessop Wing, Royal Hallamshire Hospital, Sheffield S10 2SF, UK
[2]Hull York Medical School, University of Hull, Hull, UK

*Correspondence address. E-mail: alazemimajda@hsc.edu.kw

Submitted on September 12, 2010; resubmitted on October 29, 2010; accepted on November 10, 2010

BACKGROUND: Many hormone and ultrasound measurements have been assessed as possible markers of ovarian reserve and to identify potential poor responders to ovulation induction. The objective of this study is to determine whether multiple biomarkers measured in blood samples collected immediately before commencement of ovulation induction for IVF can predict the outcome of ovarian stimulation.

METHODS: We conducted a prospective observational study, including 356 unselected women undergoing ovulation induction/IVF at two centers. Anti-Müllerian hormone (AMH), inhibin B and FSH were measured before commencement of ovulation induction. The main outcome measures were the number of oocytes retrieved and pregnancy outcome.

RESULTS: Univariate analyses showed that age, FSH, inhibin B and AMH were significant predictors for poor oocyte yield. AMH presented the highest receiver operating characteristic area under the curve (ROC_{AUC}) of 0.827 indicating a good discriminating potential for predicting poor ovarian response, followed by FSH with an ROC_{AUC} of 0.721. In the multivariate analysis, the variables age, FSH and AMH remained significant and the resulting model provided a high ROC_{AUC} of 0.819. Women with an ovarian reserve test of <0.3 have more than a 75% chance of having their treatment cycle canceled, but a value over 0.73 indicates a 38% chance of pregnancy. Number of oocytes and oocyte yield per unit FSH administered were correlated with log model for no pregnancy ($r = -0.217$, $P < 0.001$ and $r = -0.367$, $P < 0.001$, respectively) but had limited predictive value.

CONCLUSIONS: A derived estimate of ovarian reserve demonstrated superior ability for predicting oocyte yield after ovulation induction when compared with any single endocrine marker (AMH, inhibin B, FSH).

Key words: ovarian reserve / ovulation induction / anti-Müllerian hormone / inhibin B / FSH

Introduction

The success of IVF and embryo transfer is dependent on an adequate response of the ovaries to exogenous gonadotrophin stimulation (Akande *et al.*, 2002). Treatment cancelation owing to poor ovarian response is a significant problem seen in 12–30% of all stimulated cycles (Keay *et al.*, 1997). Poor ovarian response to stimulation may be a consequence of advancing chronological age (Kim, 1995) although poor response may also occur unexpectedly in relatively young patients (Nikolaou and Templeton, 2003; Lambalk *et al.*, 2009). Although neuroendocrine and uterine factors may reduce fertility with age, progressive depletion of the size of the pool of ovarian follicles is thought to be the major cause of this problem. Decline in primordial follicle number with ageing has been linked to an equivalent decline in oocyte quality with adverse factors affecting both nucleus (aneuploidy, abnormal spindle formation) and cytoplasm (reduction in mitochondrial number and ATP, abnormalities of the cytoskeleton) (Ottolenghi *et al.*, 2004). Studies controlling for male causes of infertility by the use of donor sperm (Schwartz and Mayaux, 1982) or using donated oocytes from young donors to create embryos later transferred to the endometrial cavity of older recipients (Sauer, 1998) clearly show that age-related changes in oocytes are responsible for the majority of the decline in reproductive potential in women as they approach their fifth decade of life.

There are clear advantages to identifying women at risk of a poor response to ovulation induction before commencement of an IVF treatment cycle. IVF is expensive and invasive, and should not be performed if there is not a realistic chance of pregnancy. Although clinical experience suggests that most couples will go ahead with ovulation induction even if the likely response is poor, pretreatment identification of the 'poor responder' will allow directed counseling to be

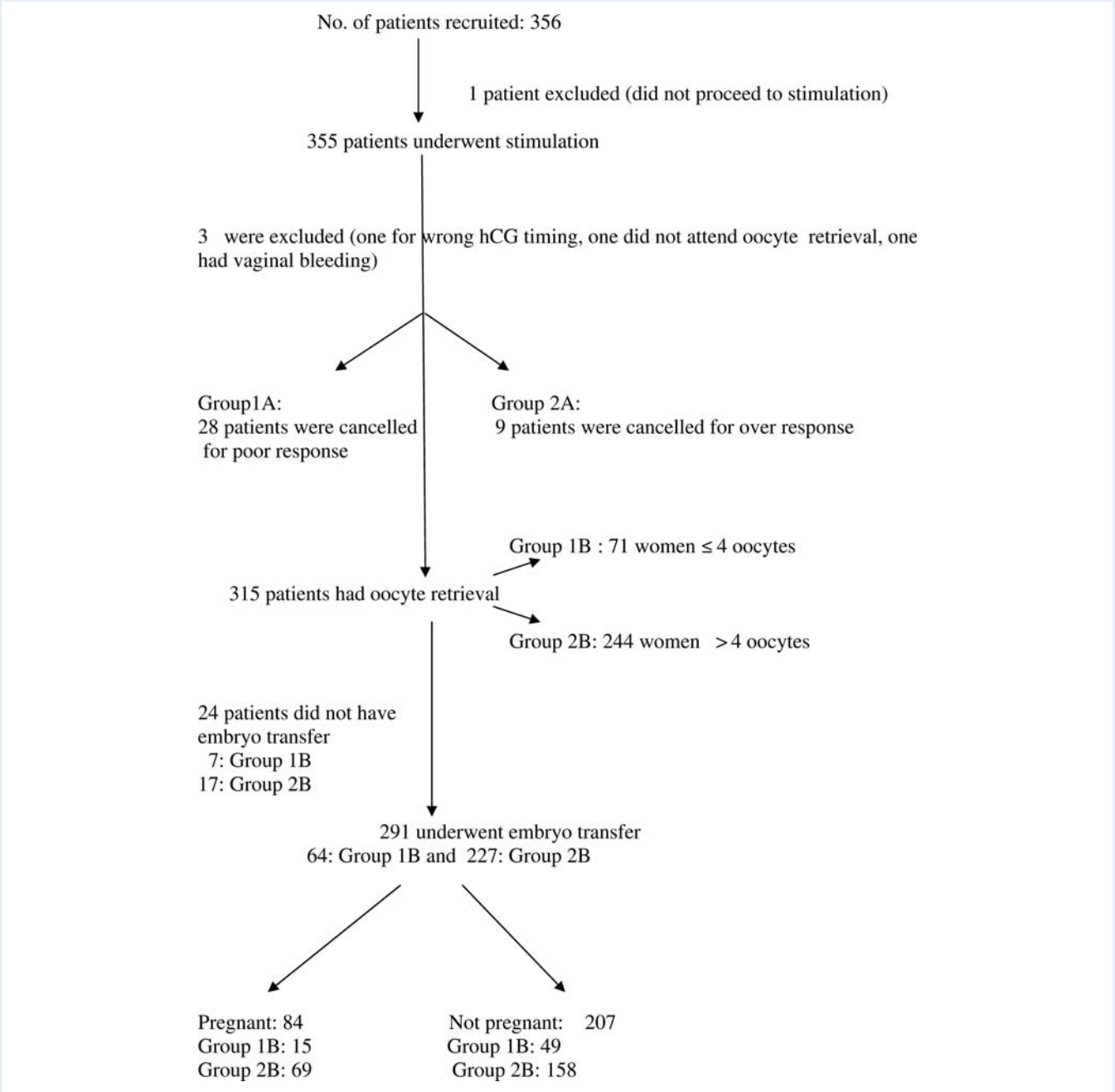

Figure 1 Flow chart showing recruitment and outcomes in women undergoing IVF.

given, lessening the disappointment of cycle cancelation or poor oocyte yield and helping couples to decide to stop treatment earlier, avoiding multiple stimulation cycles. The availability of an accurate screening test of ovarian reserve would also allow those predicted to under-respond to be given higher doses of gonadotrophin without risk of ovarian hyperstimulation syndrome (OHSS), possibly improving oocyte yield (Arslan *et al.*, 2005).

Many hormonal and ultrasound measurements have been assessed as possible markers of ovarian reserve and to identify potential poor responders to ovulation induction (Johnson *et al.*, 2006) although to date combinatorial methods have not been widely applied. Historically, combination of early follicular FSH and age was found to be better than age alone in predicting outcome in women undergoing IVF (Toner *et al.*, 1991). Many IVF centers continue to rely on early follicular phase measurement of FSH, notwithstanding limitations of month-by-month variation and disparity between assays. Inhibin B is released by granulosa cells of the developing ovarian follicle. Early follicular phase inhibin B correlates inversely with early follicular FSH level and has previously been used as an indicator of ovarian reserve (Seifer *et al.*, 1999). More recently, anti-Müllerian hormone (AMH) has

Table I Baseline, treatment and outcome characteristics of the women who underwent IVF ($n = 352$).

	Group 1A	No. of oocytes retrieved		Group 2A	*P*-value
	$n = 28$	≤4 $n = 71$	>4 $n = 244$	$n = 9$	
Baseline					
Age in years	35.8 ± 4.4	36.6 ± 3.8	33.5 ± 4.8	30.2 ± 7.0	<0.001
Cause of infertility					0.274
Unexplained	9 (32.1)	27 (38.0)	52 (21.3)	4 (44.5)	
Male factor	5 (17.9)	18 (25.4)	91 (37.3)	3 (33.3)	
Tubal	4 (14.3)	11 (15.5)	39 (16.0)	2 (22.2)	
Ovulatory	3 (10.7)	2 (2.8)	8 (3.3)	0 (0.0)	
Endometriosis	1 (3.6)	4 (5.6)	8 (3.3)	0 (0.0)	
Combined	6 (21.4)	9 (12.7)	46 (18.8)	0 (0.0)	
Treatment					
Duration of stimulation (days)	9 ± 2	9 ± 2	10 ± 2	10 ± 3	0.004
Total amount of recombinant FSH (IU/l)	1866 ± 953	2027 ± 631	2016 ± 814	1939 ± 933	0.310
Peak estradiol (pmol)	1506 ± 837	4049 ± 2385	27101 ± 5248	7371 ± 4463	<0.001
Study variables (serum)					
FSH (IU/l)	8.28 ± 2.91	8.13 ± 2.71	6.34 ± 1.89	4.41 ± 0.99	<0.001
Inhibin B (pg/ml)	28.7 ± 31.4	54.0 ± 67.5	63.1 ± 36.9	95.5 ± 36.2	<0.001
AMH (ng/ml)	2.19 ± 3.74	1.07 ± 1.08	2.64 ± 1.85	7.44 ± 3.06	<0.001

Values are mean ± SD, except for cause of infertility which are *n* (%).
P-values are generated using Kruskal–Wallis test, except for cause of infertility (χ^2 test) which is not reliable owing to small frequencies.
AMH, anti-Müllerian hormone.

emerged as another marker with relevance to ovarian function. It is produced solely by the functional granulosa cells of growing pre-antral and small antral ovarian follicles (Weenen *et al.*, 2004; Visser *et al.*, 2006), indirectly reflecting the size of the remaining primordial follicle pool and shows little inter- and intra-cycle variability (Hehenkamp *et al.*, 2006; La Marca *et al.*, 2007; Seifer and Maclaughlin, 2007; Tsepelidis *et al.*, 2007). Ultrasonographic measurements of antral follicle count (AFC) and ovarian volume have also been explored as predictors of response to ovulation induction. Whilst demonstrating good predictive ability and reliability in expert hands, problems of reproducibility limit their usefulness in practice (Broekmans *et al.*, 2006).

The presence of such a wide range of tests of ovarian reserve suggests that no single test provides a sufficiently accurate result. A test based on a combination of markers might provide better identification of diminished ovarian reserve and act as a more sensitive predictor of response to ovarian stimulation in patients undergoing IVF treatment. Kline *et al.* (2005) produced predictive models based on chronological age, ovarian volume, FSH and inhibin B. Combinations of various markers (AFC, AMH and inhibin B) have also been used to predict poor response to stimulation, with up to 87% sensitivity, 80% specificity and a positive likelihood ratio of 4.36%. However, this scoring system was not tested for prediction of pregnancy (Muttukrishna *et al.*, 2004, 2005).

We have previously demonstrated the predictive ability for cycle cancelation of a number of endocrine markers measured in a group of women at high risk of poor response to ovulation induction (McIlveen *et al.*, 2007). We now report the results of a large study on an unselected population of women entering IVF ovulation induction, powered to determine the ability of an ovarian reserve test (ORT) to predict the outcome of ovulation stimulation both in terms of oocyte yield and chance of pregnancy.

Materials and Methods

Study population

This prospective study included a total of 356 women undergoing IVF ± ICSI at the Assisted Conception Units based at the Centre for Reproductive Medicine and Fertility, Jessop Wing, Sheffield and Hull IVF Unit. Informed consent was obtained from all women and the study was approved by the South Sheffield Research Ethics Committee. All women attending for IVF treatment during the study period (January 2008–August 2009) were invited to participate, without exclusion because of previous treatments, female age, cause of infertility or other IVF related criteria. Routine practice in our unit is to restrict IVF treatment to patients with basal FSH < 13 IU/l. Therefore, there was an a priori selection bias against those with high FSH.

Of the 356 patients who agreed to participate, one withdrew before starting stimulation and three patients were canceled during stimulation for the following reasons: wrong timing of hCG (one), did not attend oocyte retrieval (one) and significant vaginal bleeding during stimulation (one). The remaining 352 women were classified into two groups.

Table II Characteristics of the studied women according to response to stimulation ($n = 352$) and pregnancy outcome ($n = 326$).

Variable	Number of oocytes		*P*-value
	Group 1 ($n = 99$)	Group 2 ($n = 253$)	
Age in years	36.3 ± 3.9	33.4 ± 4.9	<0.001
FSH (IU/l)	8.17 ± 2.75	6.28 ± 1.90	<0.001
Inhibin B (pg/ml)	46.8 ± 60.5	64.2 ± 37.4	<0.001
AMH (ng/ml)	1.39 ± 2.22	2.81 ± 2.09	<0.001
Duration of stimulation (days)	9 ± 2	10 ± 2	<0.001
Total amount of recombinant FSH (IU/l)	1980 ± 737	2013 ± 816	0.975
Peak estradiol (pmol)	3338 ± 2364	7760 ± 5243	<0.001
Number of embryos	2 ± 1	6 ± 4	<0.001
Ongoing pregnancy rate	19 (19.2)	69 (30.4)*	0.036
Pregnancy outcome	+ve ($n = 88$)	−ve ($n = 238$)	
Age in years	33.2 ± 4.6	34.9 ± 4.6	0.002
FSH (IU/l)	7.01 ± 2.32	6.91 ± 2.32	0.604
Inhibin B (pg/ml)	65.2 ± 59.7	56.2 ± 39.6	0.251
AMH (ng/ml)	2.44 ± 2.04	2.18 ± 2.02	0.038
Duration of stimulation (days)	9 ± 2	9 ± 2	0.933
Total amount of recombinant FSH (IU/l)	1783 ± 527	2095 ± 868	0.008
Peak estradiol (pmol)	5351 ± 3162	6133 ± 4498	0.588
Number of oocytes	8 ± 4	7 ± 5	0.056
Number of embryos	5 ± 3	4 ± 4	0.055

Values are mean ± SD except for ongoing pregnancy rate *n* (%).
P-values are generated by Mann–Whitney test, except for ongoing pregnancy rate (χ^2 test).
*Group 2: we excluded women at risk of hyperstimulation who did not proceed for embryo transfer (227 women did have embryo transfer).

Table III Correlation of different characteristics with number of oocytes and number of embryos in women undergoing stimulation for IVF ($n = 352$).

Variable	Number of oocytes		Number of embryos	
	r_s	*P*-value	r_s	*P*-value
Age in years	−0.286	<0.001	−0.140	0.012
FSH (IU/l)	−0.333	<0.001	−0.250	<0.001
Inhibin B (pg/ml)	0.254	<0.001	0.209	<0.001
AMH (ng/ml)	0.502	<0.001	0.378	<0.001

r_s = Spearman correlation coefficient.

(i) Group 1: Total number: 99
  (a) Group 1A: represents those who were canceled during stimulation owing to poor response and did not proceed to hCG administration and oocyte collection (28 women).
  (b) Group 1B: represents those who proceeded to oocyte retrieval and had ≤4 oocytes (71 women).
(ii) Group 2: Total number 253
  (a) Group 2A: represents those who were deemed to have an excessive response to gonadotrophins and therefore had their cycle canceled before hCG because of risk of OHSS (nine women).
  (b) Group 2B: represents those who proceeded to oocyte retrieval and had >4 oocytes (244 women).

Out of 315 who had oocyte retrieval, 24 women did not proceed to embryo transfer. Seven of these women were from group 1B and 17 were from group 2B. One patient failed to have any oocytes collected and 10 women had complete failure of fertilization. Two women had elective cryopreservation of oocytes before chemotherapy. Embryo transfer was not performed in two patients because of identification of an endometrial irregularity during stimulation (possible polyp or thin endometrium), one woman had a suspected endometritis and eight women had elective cryopreservation of all embryos because of risk of OHSS (Fig. 1).

Study protocol

All patients had blood collected on Day 2 of the menstrual cycle for measurement of FSH, LH, estradiol (E_2), inhibin B and AMH. A 4 ml blood sample was collected prior to the start of the IVF treatment cycle under investigation. Serum was stored at −20°C and analyzed later, all hormones being analyzed as a single batch. Each patient was given a unique numerical identifier which was used in data analysis to ensure anonymity of subjects. Transvaginal ultrasound was performed at the same visit using a 6-MHz probe (Toshiba, Sterling, UK). Controlled ovarian hyperstimulation was undertaken using protocols individualized according to patient age, previous response to gonadotrophins and basal FSH. Prescribing clinicians were not aware of the results of the assays for inhibin B or

Table IV Predictors for poor response in women (≤4 oocytes) and for negative pregnancy outcome using univariate and multivariate logistic regression analysis.

Factor	Odds ratio	95% CI[a]	*P*-value	ROC_{AUC}
Poor response in women				
Univariate analysis				
Age in years	1.160	1.094–1.230	<0.001	0.676
FSH (IU/l)	1.444	1.286–1.621	<0.001	0.721
Inhibin B (pg/ml)	0.987	0.979–0.994	<0.001	0.686
AMH (ng/ml)	0.562	0.450–0.703	<0.001	0.827
Multiivariate analysis				
OR log model				0.819
Age in years	1.090	1.020–1.164	0.010	
FSH (IU/l)	1.275	1.126–1.443	<0.001	
Inhibin B (pg/ml)	0.994	0.986–1.001	0.102	
AMH (ng/ml)	0.764	0.612–0.955	0.018	
Negative pregnancy outcome				
Univariate analysis				
Age in years	1.084	1.028–1.143	0.003	0.610
FSH (IU/l)	0.983	0.885–1.091	0.746	0.519
Inhibin B (pg/ml)	0.996	0.991–1.001	0.136	0.541
AMH (ng/ml)	0.942	0.840–1.057	0.310	0.575
Multivariate analysis				
Pregnancy log model				0.633
Age in years	1.096	1.033–1.163	0.002	
FSH (IU/l)	0.915	0.812–1.032	0.147	
Inhibin B (pg/ml)	0.996	0.991–1.001	0.136	
AMH (ng/ml)	0.997	0.870–1.141	0.962	

Logistic regression model: Probability of a patient to have oocytes ≤4, $P = \frac{e^{-4.800+0.086\times Age+0.243\times FSH-0.006\times InhibinB-0.269\times AMH}}{1+e^{-4.800+0.086\times Age+0.243\times FSH-0.006\times InhibinB-0.269\times AMH}}$.

Logistic regression model: Probability of a patient to have a negative pregnancy outcome, $P = \frac{e^{-1.283+0.092\times Age-0.089\times FSH-0.004\times InhibinB-0.003\times AMH}}{1+e^{-1.283+0.092\times Age-0.089\times FSH-0.004\times InhibinB-0.003\times AMH}}$.

[a]95% CI = 95% confidence interval for odds ratio. ROC_{AUC}, receiver operating characteristic area under curve.

AMH. In most cycles (248 women, 69.9%) an antagonist protocol was used, with a mid-luteal phase start GnRH agonist long protocol being used for 15.5% of patients (55 women) and a early follicular phase start GnRH agonist 'flare' protocol for 52 (14.6%) of patients.

Gonadotrophin dose was also prescribed on an individual basis according to multiple parameters including female age, Day 2 FSH, BMI and response in previous stimulation cycles. A minimum of 100 IU per day and a maximum of 450 IU per day were used. All patients received recombinant FSH without addition of LH. Dose adjustments were seldom made once the cycle had commenced.

A total of 10 000 IU of hCG were given when there were at least two leading follicles that measured ≥17 mm in the antagonist protocol and when there were at least two leading follicles that measured ≥18 mm in the GnRH agonist protocols (Devroey *et al.*, 1994, 2009). All follicles >12 mm were aspirated transvaginally under ultrasound guidance 34–36 h after hCG. Oocyte number and quality were assessed by the embryologist. Fertilization was with IVF or ICSI according to cause of infertility. A maximum of two embryos were transferred on Day 3 following oocyte collection. An ongoing pregnancy was defined as the presence of fetal cardiac activity beyond 12 weeks gestation.

Cycle cancelation was recommended if less than three follicles of ≥14 mm mean diameter were seen when the lead follicle reached 18 mm (i.e. total mature follicle count ≤2). If a couple were advised to cancel the cycle, an intrauterine insemination was offered if appropriate (patent tubes and adequate sperm concentration and motility).

Cycle cancelation was also recommended for those with excessive response, with serum E_2 > 20 000 pMol/l and/or more than 20 follicles of ≥14 mm diameter on ultrasound.

Hormone assays

FSH and E_2 were measured using an automated multi-analysis system with chemoluminescence detection (Adiva Centaur; Bayer, Newbury, UK). For FSH, functional sensitivity was 0.3 IU/l, and intra- and inter-assay variabilities were <3%. For E_2, functional sensitivity was 26 pmol/l, and intra- and inter-assay variabilities were <11 and 7%, respectively. Inhibin B concentrations were measured in duplicate using a specific two-site enzyme-linked immunosorbent assay (ELISA) kit (Oxford Bio-Innovation Ltd, Oxford, UK) according to the manufacturer's protocol. The assay detection limit for inhibin B was <5 pg/ml. Within-plate and between-plate coefficient of variation (CV) were 6 and 8%, respectively. Serum AMH

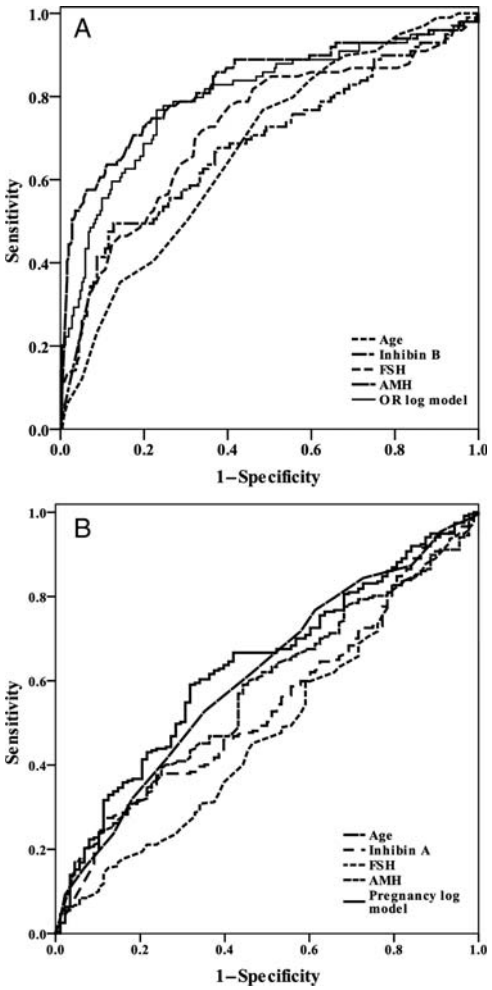

Figure 2 Receiver operating characteristic curves for predictors of: (**A**) poor response (≤4 oocytes), (**B**) negative pregnancy in women undergoing IVF. AMH, anti-Müllerian hormone; OR log model, ovarian reserve.

samples were assayed in duplicate using the AMH/MIS ELISA kit (Immunotech-Beckman, Marseilles, France) according to the manufacturer's protocol. The detection limit of the assay was <3 pmol/ml (0.42 ng/ml). Each sample was analyzed in duplicate. A standard curve was established using AMH standards that were provided in the kit, and the intra-assay CV was 12.3% and the inter-assay CV was 14.2%.

Statistical analysis

Power calculation

A sample size calculation was performed using the end-point of prediction of ongoing pregnancy using logistic regression modeling. Assuming that the blood test score will be normally distributed, the odds of ongoing pregnancy at the mean blood test score is estimated to be one-third and the odds ratio following a 1 SD reduction in blood test score from the mean is estimated to be $\sqrt{2} = 1.414$. Using a 5% significance level, this leads to a requirement of 348 participants for 80% power.

Statistical methods

We performed all statistical analyses using STATA (SE 8.2, StataCorp, College Station, TX, USA). Quantitative variables are presented as mean ± SD and qualitative variables as number (%). Normality was evaluated with the Kolmogorov–Smirnov test. Due to the non-normality of quantitative variables, Mann–Whitney test was used to compare two independent groups and Kruskal–Wallis test to compare more than two groups. The association of cause of infertility with different groups was assessed using Fisher's exact test. Spearman correlation was used to assess the association between two quantitative variables. The univariate logistic regression analyses were used to estimate the risk of each independent variable on the dependent variable and multivariate logistic regression analyses were used to assess the independent effect of these variables after controlling confounding between them. The area under the receiver operating characteristic curve (ROC_{AUC}) was computed to assess the predictive accuracy of each independent variable and the logistic regression model. We created an ORT value from multivariate logistic regression equation in which age, FSH, inhibin B and AMH were considered as predictors for poor response and negative pregnancy. A *P*-value of <0.05 was considered as statistically significant.

Results

Baseline, treatment and outcome characteristics

Baseline, treatment and outcome characteristics of the study population are summarized in Table I. Three hundred and fifty-six women were eligible for this study (Fig. 1). Twenty-eight (7.9%) treatment cycles were canceled for a poor ovarian response (Group 1A). Nine (2.5%) treatment cycles were canceled for excessive response (Group 2A). Three hundred and fifteen women underwent oocyte retrieval (315/355: 88.7%). Those who underwent oocyte retrieval were divided into two groups: Group 1B (71/352 women; 20.2%) including those with oocyte yield of ≤4 and Group 2B (244/352 women; 69.3%) including those with oocyte yield of >4. There were significant differences between different groups in terms of age, duration of stimulation, peak E_2 level as well as FSH, inhibin B and AMH but not in the cause of infertility (Table I). Table II compared Group 1 (1A and 1B) with Group 2 (2A and 2B). Patients in Group 1 were older, having lower concentrations of inhibin B and AMH and higher FSH concentrations compared with those in Group 2. Group 2 had significantly higher number of embryos as well as a higher ongoing pregnancy rate compared with Group 1.

Twenty-six women (Group 2A who were canceled prior to oocyte retrieval because of an excessive response and 17 women who were had their embryos frozen for different reasons) were excluded from the pregnancy analysis. Group 1A were included in the pregnancy analysis as their poor response was the reason for their cancelation. Comparison between those with ongoing pregnancy (88, 27%) and those without (238, 73%) revealed that there were significant differences in age ($P = 0.002$), AMH ($P = 0.038$) and total amount of recombinant FSH used ($P = 0.008$) between the two groups (Table II). Women with ongoing pregnancy were younger, had higher AMH and received lower amounts of recombinant FSH. There were borderline significant differences in number of oocytes and number of embryos between the two groups ($P = 0.056$ and $P = 0.055$, respectively). No significant differences were observed in concentrations of FSH or inhibin B, duration of stimulation or peak E_2 level.

Table V Comparison of performance characteristics for poor responders (≤4 oocytes) and for negative pregnancy outcome.

Variable	Cutoff value	Sensitivity (%)	Specificity (%)	Correctly classified (%)	LR+
Poor responders					
Age in years	36	63.6	60.5	61.4	1.61
FSH (IU/l)	7.0	69.7	67.9	68.4	2.17
Inhibin B (pg/ml)	49.4	64.0	63.6	63.9	1.76
AMH (ng/ml)	1.36	75.5	74.8	75.3	2.99
OR log model	0.30	76.8	76.6	76.6	3.28
Negative pregnancy outcome					
Age in years	35	61.8	53.4	59.5	1.33
FSH (IU/l)	6.8	53.4	52.4	52.6	1.12
Inhibin B (pg/ml)	53.2	50.0	49.6	49.7	0.99
AMH (ng/ml)	1.76	56.8	56.3	56.4	1.30
Pregnancy log model	0.73	62.5	61.4	62.2	1.62

LR+, likelihood ratio of a positive test = Sensitivity/(1 − Specificity).

Table III shows that there were significant correlations between different variables (age, FSH, inhibin B and AMH) for number of oocytes as well as number of embryos. The correlation between AMH and number of oocytes was the strongest ($r_s = 0.502$).

The results of the logistic regression analyses for the prediction of poor response (≤4 oocytes) are given in Table IV. Univariate analyses showed that age, FSH, inhibin B and AMH were significant predictors for poor oocyte yield. AMH presented the highest ROC_{AUC} of 0.827, indicating a good discriminating potential for predicting poor ovarian response, followed by FSH with an ROC_{AUC} of 0.721. In the multivariate analysis, the variables age, FSH and AMH remained significant and the resulting model provided a high ROC_{AUC} of 0.819.

ROC curve analysis facilitates the selection of cutoff values that maximize sensitivity, specificity and accuracy (Fig. 2). The ovarian reserve log model ROC curve for poor responders estimated that an ORT value of 0.30 produced a maximized sensitivity of 76.8% and a specificity of 76.6%. (Table V). The results of univariate and multivariate logistic regression analyses showed that age was the only significant predictor for negative pregnancy outcome with an ROC_{AUC} of 0.610 (Table IV). None of the individual markers of ovarian reserve were able to predict pregnancy. However, adding these endocrine markers to the pregnancy log model will increase ROC by only 0.023. The pregnancy log model ROC curve for negative pregnancy outcome estimated that a value of 0.73 produced a maximized sensitivity of 62.5% and a specificity of 61.4% (Table V). Women with an ORT of <0.3 have more than a 75% chance of having their treatment cycle canceled, but a value over 0.73 indicates a 38% chance of pregnancy. Figure 3 shows the relationship between the log model values for no pregnancy and Fig. 3A number of oocytes; Fig. 3B oocyte yield per unit FSH dose. Both number of oocytes and oocyte yield per unit FSH dose were significantly correlated with log model for no pregnancy ($r = -0.217$, $P < 0.001$ and $r = -0.367$, $P < 0.001$, respectively).

Discussion

This prospective study demonstrated that a derived multi-marker computation for measuring ORT was a good predictor for oocyte yield after ovulation induction and also for ongoing pregnancy. The analysis showed that the ability of the derived ORT to predict pregnancy was related to the inclusion of age in the ORT formula. The ability of the test to predict pregnancy was relatively modest, which is unsurprising given the large number of other variables that impact on chances of ongoing pregnancy after IVF and which are external to ovarian reserve. These include the effects of sperm quality, endometrial receptivity and other uterine factors. Both endocrine and ultrasound methods of assessing ovarian reserve appear to measure the quantity of follicles in the developing cohort rather than oocyte 'quality' (Broekmans *et al.*, 2006). In a subfertile population, the ideal test would accurately predict oocyte quality as well as oocyte yield, with better ability to determine the likelihood of pregnancy and live birth. ORT had the ability to predict chance of ongoing pregnancy.

The ability of the computed ORT to predict pregnancy outcome is further enhanced by inclusion of the oocyte yield per unit FSH dose, as shown in Fig. 3. This computation takes into account the finding that pregnancy is less likely in those women who have poor oocyte yield notwithstanding treatment with high doses of FSH, representing the most gonadotrophin-resistant group of 'poor responders'. Several hormonal markers and ultrasound parameters have been used to estimate ovarian reserve and predict those with a poor chance of success in assisted reproduction techniques: these include age, FSH, LH, E_2, inhibin B, AMH, ovarian volume and ovarian AFC. Others have advocated the use of dynamic tests using ovarian response to GnRH, FSH or clomiphene citrate to assess ovarian function (Coccia and Rizzello, 2008).

AMH, produced by granulosa cells of pre-antral and small antral follicles, has emerged as a useful marker of ovarian function. AMH

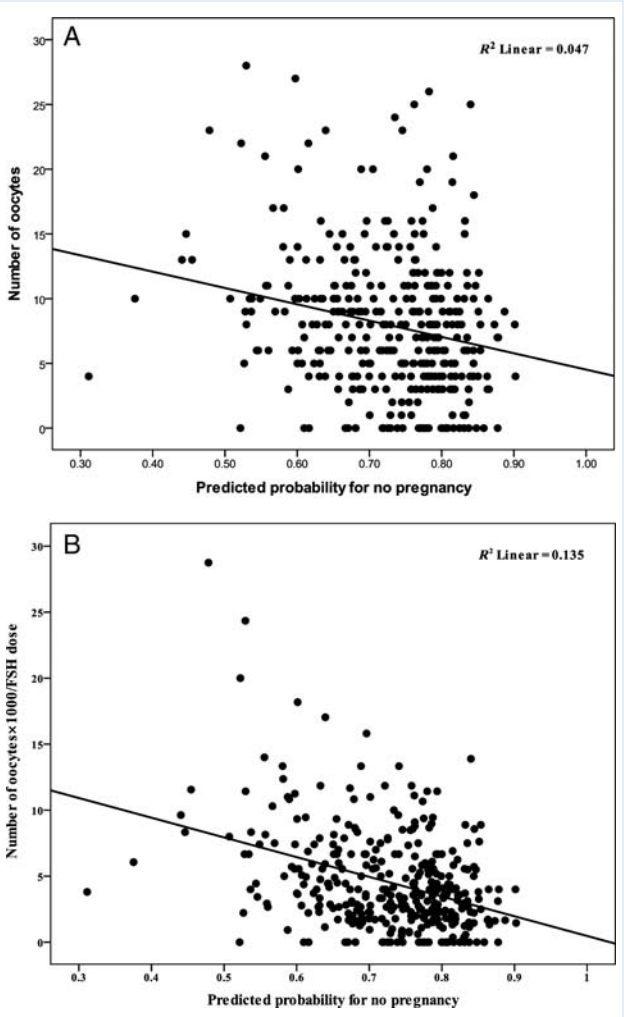

Figure 3 Relationship between predicted probability for negative pregnancy and (**A**) number of oocytes, (**B**) number of oocytes/FSH dose for women undergoing IVF.

has been used in assessment of ovarian aging, prediction of response to ovulation induction and the assessment of the risk of developing OHSS (Van Rooij *et al.*, 2002; Nelson *et al.*, 2007). Our study clearly demonstrated the superiority of AMH in prediction of poor response compared with the other individual markers (AUC for inhibin B: 0.686; FSH: 0.721; AMH: 0.827). Nelson *et al.* (2009) investigated the role of AMH in predicting oocyte yield, showing that the use of circulating AMH concentration to individualize treatment strategies for controlled ovarian stimulation reduced clinical risk of OHSS whilst optimizing pregnancy rates. Our model included AMH with results from Day 2 FSH, inhibin B and patient age to derive the ORT. Our study has shown that ORT is equivalent to AMH in predicting oocyte yield (AUC for AMH: 0.827 and for ORT: 0.819) but superior in predicting pregnancy outcome. This is mainly a result of the inclusion of female age in the ORT formula. This model could be used clinically for testing of ovarian reserve and prediction of oocyte yield. Further refinement is needed before chance of pregnancy can be better predicted with sufficient accuracy for clinical use, probably with inclusion of novel determinants of oocyte quality, sperm function and endometrial receptivity. It will also be of interest to study the summative chance of pregnancy once the patients included in the study have completed transfer of their frozen embryos. These data will take several years to accrue but it seems likely that the ability of the ORT to predict pregnancy will improve since those who score highest in the test have the greatest chance of having frozen embryos for use later.

Not all women will respond normally to ovulation induction. Use of biochemical markers measured before commencement of stimulation has been shown to identify those at high risk of poor response (McIlveen *et al.*, 2007), findings confirmed in the deliberately heterogeneous group of patients assessed in our study. Forewarning of poor response will reduce the stress caused by cycle cancelation or low oocyte yield and may help the couple to move on to other more appropriate treatments for their infertility and avoid the cost and futility of repeated stimulation cycles. Equally, pretreatment assessment of ovarian reserve can be used to identify patients at risk of ovarian hyperstimulation (Nelson *et al.*, 2009). Such patients require treatment with low doses of FSH in a GnRH antagonist controlled cycle, with careful monitoring and a low threshold for cycle cancelation or cryopreservation of all embryos without fresh transfer (Mathur *et al.*, 2007). Moreover, the use of GnRH agonist for triggering the final oocyte maturation in a GnRH antagonist controlled cycle is now a valid preventive measure for high-risk patients (Humaidan *et al.*, 2010).

An accelerated depletion of primordial follicles appears to occur around the age of 38 years (Fady and Gosden, 1995). The onset of the effects of diminished ovarian reserve on fertility varies from individual to individual and can be detected at a younger age in some women (Scott and Hofmann, 1995). The accelerated diminution of ovarian reserve is reflected by a rise in serum concentration of FSH and reduced concentrations of AMH. Young women with poor ovarian response to ovulation induction during IVF treatment are at risk of early menopause (Nikolaou *et al.*, 2002) confirming that ovarian response to ovulation induction can be used as a surrogate marker for overall ovarian reserve.

Hence, we can use the number of oocytes recovered after ovulation induction as a surrogate marker for natural ovarian reserve. Further research to test the ability of ORT to identify individuals with a diminished (or possibly also enhanced) ovarian reserve in the general, non-infertile population is needed, although these studies will of necessity be very long-term. In recent years, the tendency to postpone childbearing (Astolfi *et al.*, 1999) has resulted in more and more Western women trying to have children at a more advanced age, with increased likelihood of 'unexplained' infertility (Van Zonneveld *et al.*, 2001). Screening for 'early ovarian aging' in the late 20s or early 30s using ORT could provide information to women to allow them to make rational decisions about their fertility, allowing those with a smaller than average follicle pool to consider early attempts to conceive, or perhaps to decide to cryopreserve oocytes or embryos for use later.

Acknowledgements

We thank the staff of the Assisted Conception Units at Jessop Wing, Royal Hallamshire Hospital Sheffield and Hull IVF Unit for their assistance, and all the patients who participated in the study.

References

Akande VA, Fleming CF, Hunt LP, Keay SD, Jenkins JM. Biological versus chronological ageing of oocytes, distinguishable by raised FSH levels in relation to the success of IVF treatment. *Hum Reprod* 2002;**17**:2003–2008.

Arslan M, Bocca S, Mirkin S, Barroso G, Stadtmauer L, Oehninger S. Controlled ovarian hyperstimulation protocols for in vitro fertilization: two decades of experience after the birth of Elizabeth Carr. *Fertil Steril* 2005;**84**:555–569.

Astolfi P, Ulizzi L, Zonta LA. Selective cost of delayed childbearing. *Hum Reprod* 1999;**14**:572–573.

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.

Coccia ME, Rizzello F. Ovarian reserve. *Ann N Y Acad Sci* 2008; **1127**:27–30 (Review).

Devroey P, Mannaerts B, Smitz J, Coelingh Bennink H, Van Steirteghem A. Clinical outcome of a pilot efficacy study on recombinant human follicle-stimulating hormone (Org 32489) combined with various gonadotrophin-releasing hormone agonist regimens. *Hum Reprod* 1994;**9**:1064–1069.

Devroey P, Aboulghar M, Garcia-Velasco J, Griesinger G, Humaidan P, Kolibianakis E, Ledger W, Tomás C, Fauser BC. Improving the patient's experience of IVF/ICSI: a proposal for an ovarian stimulation protocol with GnRH antagonist co-treatment. *Hum Reprod* 2009;**24**:764–774.

Faddy MJ, Gosden RG. A mathematical model of follicle dynamics in the human ovary. *Hum Reprod* 1995;**10**:770–775.

Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, Te Velde ER, Broekmans FJ. Anti-Müllerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. *J Clin Endocrinol Metab* 2006;**91**:4057–4063.

Humaidan P, Quartarolo J, Papanikolaou EG. Preventing ovarian hyperstimulation syndrome: guidance for the clinician. *Fertil Steril* 2010 April 21 (Epub ahead of print).

Johnson NP, Bagrie EM, Coomarasamy A, Bhattacharya S, Shelling AN, Jessop S, Farquhar C, Khan KS. Ovarian reserve tests for predicting fertility outcomes for assisted reproductive technology: the International Systematic Collaboration of Ovarian Reserve Evaluation protocol for a systematic review of ovarian reserve test accuracy. *BJOG* 2006;**113**:1472–1480.

Keay SD, Liversedge NH, Mathur RS, Jenkins JM. Assisted conception following poor ovarian response to gonadotrophin stimulation. *Br J Obstet Gynaecol* 1997;**104**:521–527.

Kim SH. Female aging and superovulation induction for IVF. *J Obstet Gynaecol* 1995;**21**:75–82.

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

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

Lambalk CB, van Disseldorp J, de Koning CH, Broekmans FJ. Testing ovarian reserve to predict age at menopause. *Maturitas* 2009;**63**:280–291.

Mathur R, Kailasam C, Jenkins J. Review of the evidence base of strategies to prevent ovarian hyperstimulation syndrome. *Hum Fertil (Camb)* 2007; **10**:75–85.

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

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

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

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

Nelson SM, Yates RW, Lyall H, Jamieson M, Traynor I, Gaudoin M, Mitchell P, Ambrose P, Fleming R. Anti-Müllerian hormone-based approach to controlled ovarian stimulation for assisted conception. *Hum Reprod* 2009;**24**:867–875.

Nikolaou D, Templeton A. Early ovarian ageing: a hypothesis. Detection and clinical relevance. *Hum Reprod* 2003;**18**:1137–1139.

Nikolaou D, Lavery S, Turner C, Margara R, Trew G. Is there a link between an extremely poor response to ovarian hyperstimulation and early ovarian failure? *Hum Reprod* 2002;**17**:1106–1111.

Ottolenghi C, Uda M, Hamatani T, Crisponi L, Garcia JE, Ko M, Pilia G, Sforza C, Schlessinger D, Forabosco A. Aging of oocyte, ovary, and human reproduction. *Ann N Y Acad Sci* 2004;**1034**:117–131.

Sauer MV. The impact of age on reproductive potential: lessons learned from oocyte donation. *Maturitas* 1998;**30**:221–225.

Schwartz D, Mayaux MJ. Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS. *N Engl J Med* 1982;**306**:404–406.

Scott RT Jr, Hofmann GE. Prognostic assessment of ovarian reserve. *Fertil Steril* 1995;**63**:1–11.

Seifer DB, Maclaughlin DT. Mullerian inhibiting substance is an ovarian growth factor of emerging clinical significance. *Fertil Steril* 2007; **88**:539–546.

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

Toner JP, Philput CB, Jones GS, Muasher SJ. Basal follicle-stimulating hormone level is a better predictor of in vitro fertilization performance than age. *Fertil Steril* 1991;**55**:784–791.

Tsepelidis S, Devreker F, Demeestere I, Flahaut A, Gervy Ch, Englert Y. Stable serum levels of anti-Müllerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. *Hum Reprod* 2007;**22**:1837–1840.

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

Van Zonneveld P, Scheffer GJ, Broekmans FJ, te Velde ER. Hormones and reproductive aging. *Maturitas* 2001;**38**:83–91.

Visser JA, de Jong FH, Laven JS, Themmen AP. Anti-Müllerian hormone: a new marker for ovarian function. *Reproduction* 2006;**131**:1–9.

Weenen C, Laven JS, Von Bergh AR, Cranfield M, Groome NP, Visser JA, Kramer P, Fauser BC, Themmen AP. Anti-Müllerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum Reprod* 2004; **10**:77–83.