Human Reproduction Vol.22, No.3 pp. 778–785, 2007 doi:10.1093/humrep/del435
Advance Access publication November 17, 2006.

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

M.McIlveen[1,2,4], J.D.Skull[1] and W.L.Ledger[1,3]

[1]Assisted Conception Unit, Sheffield, UK, [2]Department of Obstetrics and Gynaecology, University of Sydney, Sydney, NSW, Australia and [3]Academic Unit for Reproductive and Developmental Medicine, The University of Sheffield, Sheffield, UK

[4]To whom correspondence should be addressed at: Department of Obstetrics and Gynaecology, University of Sydney, Sydney, NSW, Australia. E-mail: myvanwy@mcilveen.com.au

BACKGROUND: Unexpectedly poor response leading to IVF cycle cancellation is a distressing treatment outcome. We have prospectively assessed several markers of ovarian reserve in a high risk IVF population to determine their utility in predicting IVF cycle cancellation. METHODS: Eighty-four women at high risk of cycle cancellation due to raised FSH, previous poor response and/or age ≥40 years attending for high-dose short protocol IVF treatment had baseline measures of FSH, inhibin B, anti-Müllerian hormone (AMH), antral follicle count (AFC) and ovarian volume. A GnRH agonist was then administered and, 24 h later, estradiol (E_2) and inhibin B measures were repeated. RESULTS: Fifty-seven per cent of patients in this study had a poor response to stimulation, and 15% were cancelled. Using multivariate logistic regression, we found that day 3 inhibin B levels were the best predictor of cycle cancellation with an area under the receiver operating curve (ROC AUC) = 0.78 (P = 0.017). When only considering baseline variables, mean ovarian volume was the best predictor of cycle cancellation (ROC AUC = 0.78; P = 0.016). AMH concentrations were the best predictor of a poor response (P = 0.003), and AMH was also predictive of cycle cancellation (P = 0.007) with very little inter-cycle variability. None of the parameters studied were predictive of ongoing pregnancy. CONCLUSIONS: This group of at-risk patients had a high rate of poor response to simulation and cancellation. Although several measures of ovarian reserve were able to predict cycle cancellation, none were able to predict pregnancy. AMH was predictive of both cycle cancellation and poor response with little inter-cycle variability.

Key words: anti-Müllerian hormone/cycle cancellation/IVF/inhibin B/ovarian reserve

Introduction

The accurate determination of ovarian reserve continues to be a challenge for reproductive physicians. In the context of IVF treatment, ovarian reserve testing can be predictive of both response to gonadotrophin stimulation and chances of success with treatment (Pellicer *et al*., 1987; Jenkins *et al*., 1991). The various available measures of ovarian reserve are better at predicting response to gonadotrophin stimulation than the chance of pregnancy (Creus *et al*., 2000; Hendriks *et al*., 2005). Rather than concentrating on attempting to predict pregnancy, it may be more helpful to use ovarian reserve testing in the prediction of cycle cancellation and poor response to stimulation. Avoiding gonadotrophin treatment for women destined not to respond to stimulation would help to reduce cancellation rates, treatment costs and emotional stress for the patient. Pretreatment counselling for predicted poor responders may ameliorate subsequent disappointment and distress.

Age, whether or not combined with basal FSH measurement, is only moderately successful as a predictor of response to superovulation, especially in women <40 years of age (Toner *et al*., 1991; Pearlstone *et al*., 1992; Scott *et al*., 1995). In our study, we used three simple factors (age, basal FSH levels and previous poor response) to identify a group of patients at risk of cycle cancellation.

Inhibin B is a dimeric polypeptide produced by ovarian granulosa cells. Anti-Müllerian hormone (AMH) is a glycoprotein hormone that, like inhibin B, is produced by granulosa cells in the adult ovary (Vigier *et al*., 1984). When measured basally, inhibin B and AMH concentrations can reflect the number of follicles that will ultimately reach maturity (Seifer *et al*., 1997, 2002; Tinkanen *et al*., 1999; te Velde and Pearson, 2002; Muttukrishna *et al*., 2004). Other studies have also found that inhibin B and AMH concentrations decline before a rise in basal FSH levels and thus are earlier markers of a reduction in ovarian reserve (Seifer *et al*., 1999; de Vet *et al*.,

2002). AMH may be a better marker of ovarian responsiveness than inhibin B, as it may reflect the size of the larger resting pool of pre-FSH-dependent follicles (Fanchin *et al*., 2003; Muttukrishna *et al*., 2004). AMH appears to have less inter-cycle variability than other markers of ovarian reserve (Fanchin *et al*., 2005).

The ovarian response to GnRH agonist (GnRHa) administration (the GAST test) is a dynamic test of ovarian reserve. A 2-fold rise or more in estradiol (E_2) concentration in response to GnRHa has been shown to be predictive of IVF success (Padilla *et al*., 1990; Winslow *et al*., 1991; Ranieri *et al*., 1998). The GAST test can also measure dynamic inhibin B response. Measuring the rise in inhibin B after GnRHa administration was found to be better than age and basal FSH in predicting IVF response in a group of unselected patients (Ravhon *et al*., 2000).

The exogenous FSH ovarian reserve test (EFORT) is an alternative well-validated dynamic measure of ovarian reserve (Fanchin *et al*., 1994). Attempts to incorporate inhibin B testing into this EFORT model have also been studied and appear to be more predictive of poor response than basal levels (Dzik *et al*., 2000; Eldar-Geva *et al*., 2000; Yong *et al*., 2003). However, we chose to use the GAST because recombinant FSH (rFSH) is a more expensive alternative.

Measurement of the number of antral follicles by high-resolution ultrasound scanning has been shown to be predictive of ovarian response in a number of studies (Bancsi *et al*., 2002; Hendriks *et al*., 2005). In a recent meta-analysis, an antral follicle count (AFC) was a better predictor of ovarian reserve than a basal FSH level (Hendriks *et al*., 2005). Similarly, ovarian volume measurements have been used to estimate ovarian reserve (Lass *et al*., 1997; Tomas *et al*., 1997). In women with ovarian volumes <3 ml, the risk of cycle cancellation is increased (Lass *et al*., 1997).

The optimum protocol for ovarian stimulation in women likely to be poor responders to gonadotrophins is unclear. Several approaches have been tried to improve ovarian response and oocyte yield, but none has emerged as superior (Akman *et al*., 2001; Weissman *et al*., 2003; Cheung *et al*., 2005). In the present study, we tried to minimize variation in patient response due to variation in stimulation protocol by using a standardized regime for superovulation for all patients, with a fixed dose of rFSH. Utilizing a short protocol allowed us to study measures of ovarian reserve within the same cycle as stimulation and IVF treatment occurred.

Materials and methods

Study population

Eighty-four women (age 26–44 years) undergoing IVF ± ICSI at the Assisted Conception Unit, The Jessop Wing, Sheffield, UK, were studied prospectively between March 2004 and June 2005. To be included, patients had to meet one or more of the following criteria: (i) have a previously raised early follicular phase FSH concentration >10 IU/l, (ii) be >39 years of age and (iii) have had a previous poor response to stimulation during an IVF cycle. A previous poor response was defined as the collection of four or less oocytes or cycle cancellation. Women with only one ovary were excluded from the study. Cycle cancellation was defined as the failure to reach oocyte collection. Informed consent was obtained from all women. This study was approved by the South Sheffield Research Ethics Committee.

Twenty-three of the women also had measurements taken in a subsequent cycle to compare the inter-cycle variability of the different measures.

Study protocol

All patients had blood collected on day 2 of the menstrual cycle for measurement of serum concentrations of E_2, FSH, inhibin B and AMH. Transvaginal ultrasound was performed at the same visit using a 6-MHz probe (Toshiba, Sterling, UK). AFC and ovarian volume measurements were performed. AFC was defined as the total number of antral follicles sized 2–10 mm counted in both ovaries. Ovarian volume was calculated using the formula for the volume of an ellipsoid ($\pi/6 \times$ length $\times$ width $\times$ height). The mean volume was then determined. Ovaries with cysts >15 mm present were excluded from the analysis of ovarian volume. Buserelin acetate (Aventis Pharma Ltd, West Malling, Kent, UK) 0.5 mg was then administered s.c.

Twenty-four hours later (cycle day 3), a second blood sample was collected to measure E_2 and inhibin B. The rise in E_2 concentration was calculated by dividing day 3 E_2 concentrations by day 2 E_2 concentrations to obtain a ratio, the GAST test (Winslow *et al*., 1991). The difference in E_2 concentrations between days 3 and 2 (ΔE_2) was calculated by subtracting the day 2 from the day 3 level (Padilla *et al*., 1990).

Following the second venepuncture, buserelin (0.5 mg) was continued as a daily injection, and daily injections of rFSH 250 IU (Puregon®; Organon, Cambridge, UK) were commenced in a short GnRHa protocol. When three or more follicles ≥17 mm were observed, hCG (Pregnyl; Organon, Cambridge, UK) 10 000 IU was given s.c. and transvaginal oocyte recovery was performed ~36 h later. Embryo transfer was performed 2 or 3 days following oocyte recovery. Luteal support was provided with progesterone pessaries (Cyclogest, Alphapharma, Barnstaple, UK) 400 mg/day.

Cycle cancellation was recommended if two or less subsidiary follicles of ≥14 mm mean diameter were seen when the lead follicle reached 18 mm mean diameter (i.e. total mature follicle count <4). Some couples chose to proceed with oocyte collection and IVF against medical advice. If a couple were advised to cancel the cycle, an intrauterine insemination (IUI) was offered if appropriate (patent tubes and adequate sperm concentration and motility). Cycle cancellation was defined as those women who did not proceed with oocyte collection. An ongoing pregnancy was defined as the presence of fetal cardiac activity beyond 12 weeks' gestation.

Hormonal assays

FSH and E_2 concentrations were determined using an automated multi-analysis system with chemiluminescence 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 samples were assayed in duplicate using a commercial enzyme-linked immunosorbent assay (ELISA) kit (Oxford Bio-Innovation Ltd, Oxford, UK) according to the manufacturer's protocol. The sensitivity of the assay was 15 pg/ml. The intra- and inter-assay variabilities were <7%. AMH samples were assayed in duplicate using a commercial ELISA kit (Immunotech, Beckman Coulter UK Ltd, High Wycombe, Buckinghamshire, UK) according to the manufacturer's protocol. The sensitivity of the assay was 0.24 ng/ml. The intra- and inter-assay variabilities were <5 and 8%, respectively.

Statistical analysis

Two power calculations were performed to determine the sample size required for both univariate and multivariate logistic regression models (with cycle cancellation as the binary response variable). The estimated sample size required for univariate analyses to achieve 80% power to detect an odds ratio of 0.40, assuming one normally distributed covariate in the model and $\alpha = 0.050$ (two-sided), was 58. For a model with more than one covariate (referred to as mulitvariate), assuming the same parameters as the univariate model, 73 subjects would be required to achieve 80% power.

Data were analysed with the Statistical Package for the Social Sciences (SPSS Inc., Chicago, USA) and the Medcalc (Medcalc Software, Mariakerke, Belgium) programs. The Student's *t*-test was used to compare mean baseline variables. The area under the receiver operating characteristic curve (ROC AUC) was computed to assess the predictive accuracy of the various tests, yielding values from 0.5 (no predictive power) to 1.0 (perfect prediction). Using the results of the ROC analysis, we defined an appropriate threshold level for each test and determined the sensitivity and specificity of that threshold. The chi-square for that threshold against several outcomes (cycle cancellation, poor response and pregnancy) was then determined. The likelihood ratio (LR) was calculated for that threshold.

Logistic regression analysis was applied to study the value of age and study variables for the prediction of cycle cancellation. Multiple logistic regression analysis with forward selection of parameters was applied with $P < 0.10$ for entry.

The paired *t*-test was used to compare the inter-cycle variability of the parameters. A P value <0.05 was considered significant.

Results

Baseline, treatment and outcome characteristics

Thirteen of the 84 patients (15%) had their cycle cancelled because of a poor response. Sixty-eight patients had an embryo transfer (81%). Forty-eight patients (57%) had a poor response defined as the collection of four oocytes or less or cycle cancellation.

Baseline, treatment, cycle and study variable characteristics are summarized in Table I. The patients who were cancelled did not differ significantly in their age or in the duration, type or cause of infertility compared with those that were not cancelled. They did differ in their previously highest recorded FSH; women who were cancelled had a higher mean FSH level previously recorded ($P = 0.001$). Of note, mean FSH concentrations were high in all groups, with the overall mean being 10 IU/l.

All patients received a similar duration and amount of rFSH. There were no differences in the duration of stimulation or the amount of rFSH used between those who were and were not cancelled.

Cycle outcomes are summarized for the patients who had an oocyte recovery in Table I. The mean number of oocytes

Table I. Baseline, treatment and outcome characteristics

Variable	Overall group (n = 84)	Cycle cancelled (n = 13)	Ooocyte collection (TVOPU) performed (n = 71)	P[a]
Baseline				
Age (years ± SD)	37.3 ± 3.9	37.5 ± 5.0	37.3 ± 3.8	0.90
Primary infertility [*n* (%)]	57 (68)	9 (69)	48 (68)	0.43
Cause of infertility [*n* (%)]				
Unexplained	30 (36)	8 (61)	22 (31)	
Male factor	32 (38)	4 (31)	28 (39)	
Tubal	13 (15)	0 (0)	13 (10)	
Endometriosis	9 (11)	1 (8)	8 (11)	0.17
Duration of infertility (years ± SD)	4.6 ± 3.7	4.3 ± 5.1	4.6 ± 3.4	0.74
Highest FSH level recorded (IU/l) (range)	10.0 ± 2.9 (4.5–17.4)	12.3 ± 2.5 (7.1–16.9)	9.6 ± 2.7 (4.5–17.4)	0.001
Treatment				
Duration of stimulation (days ± SD)	9.3 ± 2.0	8.9 ± 2.5	9.4 ± 1.9	0.48
Total amount recombinant FSH [intrauterine insemination (IUI) ± SD]	2321 ± 496	2231 ± 633	2334 ± 471	0.48
Peak estradiol (E_2) level (pmol/l ± SD)	5497 ± 3148	NA	5497 ± 3148	
Outcome				
Number of oocytes (± SD)	4.7 ± 2.0	NA	4.7 ± 2.0	
Fertilization rate (%)	67 ± 25	NA	67 ± 25	
Number of embryos transferred (± SD)	1.9 ± 0.4	NA	1.9 ± 0.4	
Ongoing pregnancy rate (%)	9 (11)	1 (8)[b]	8 (11)	
Study variables				
Baseline FSH level (IU/l) (range)	8.3 ± 2.3 (2.6–14.2)	9.2 ± 2.2 (5.0–12.0)	8.1 ± 2.3 (2.6–14.2)	0.13
Baseline E_2 level (pmol/l)	167 ± 62	145 ± 54	171 ± 62	0.18
E_2 level day 3 (pmol/l)	405 ± 179	303 ± 96	424 ± 185	0.03
Inhibin B baseline (pg/ml)	58 ± 55	28 ± 43	64 ± 55	0.03
Inhibin day 3 (pg/ml)	116 ± 88	50 ± 58	129 ±87	0.002
Anti-Müllerian hormone (ng/ml)	1.6 ± 0.92	0.93 ± 0.6	1.72 ± 0.92	0.004
Antral follicle count	7.4 ± 3.5	5.2 ± 2.5	7.8 ± 3.5	0.02
Mean ovarian volume (ml)	8.0 ± 3.9	5.1 ± 2.5	8.4 ± 3.9	0.01

NA, not applicable.
Values are presented as mean ± SD or %.
[a]*P* values are for comparison between cancelled and not cancelled patients.
[b]Pregnancy from IUI treatment.

Table II. Receiver operating characteristic curve analysis of study variables for the prediction of IVF cycle cancellation expressed as sensitivity, specificity and likelihood ratios (LRs)

Variable	Area under the receiver operating characteristic curve	Threshold	Sensitivity	Specificity	LR [95% confidence interval (95% CI)]	*P*
Baseline FSH level (IU/l)	0.64	≥10	46	79	2.2 (1.0–4.6)	0.06
GAST estradiol (E_2) level day 3 (pmol/l)	0.70	≤355	77	59	1.9 (1.3–2.9)	0.04
Ratio E_2 rise days 2–3	0.60	≤2.0	38	66	1.1 (0.5–2.4)	0.65
ΔE_2 days 2–3 (pmol/l)	0.69	≤195	77	59	1.9 (1.3–2.9)	0.04
Inhibin B baseline (pg/ml)	0.67	≤30	77	69	2.5 (1.6–3.9)	0.002
Day 3 (pg/ml)	0.78	≤60	69	79	3.3 (1.8–6.0)	0.001
Sum (pg/ml)	0.77	≤65	62	80	3.1 (1.6–6.1)	0.002
Anti-Müllerian hormone (ng/ml)	0.78	≤1.25	85	63	2.3 (1.6–3.4)	0.001
Antral follicle count	0.74	≤5	46	80	2.3 (1.0–4.9)	0.06
Mean ovarian volume (ml)	0.78	≤5.3	80	75	3.3 (1.9–5.5)	0.001

GAST, test of ovarian response to GnRH agonist (GnRHa) administration.

collected was 4.7. The fertilization rate in this study (67%) was equivalent to that of the clinic average for this period. Pregnancy rates were significantly lower than the clinic's mean for this period of 35% per cycle started.

Ovarian reserve tests

The mean levels of the different study variables between cancelled cycles and those that are not cancelled are summarized in Table I. The two groups varied significantly for most of the factors studied. Only baseline FSH and E_2 concentrations did not differ between the two groups.

Table II summarizes ROC AUC, sensitivity, specificity, LRs and statistical significance for the ovarian reserve markers studied in relation to cycle cancellation. The variables that were predictive of cycle cancellation included the day 3 E_2 level, inhibin B (both days 2 and 3 concentrations), AMH and mean ovarian volume. Baseline FSH, ratio of E_2 rise (GAST) and the AFC were not predictive of cycle cancellation (Table II).

Univariate logistic regression analysis showed that several variables were good predictors of cycle cancellation similar to the ROC AUC LR analysis (Table II). Age, baseline FSH and the ratio of E_2 rise were not predictive of cycle cancellation. Although the AFC was not predictive of cycle cancellation using LR analysis, it was predictive in the univariate logistic regression analysis. For the multivariate logistic regression analysis using stepwise forward selection on all variables summarized in Table III, the inhibin B level on day 3 was selected in the first step, and no other variable contributed significantly thereafter ($P = 0.017$).

Dynamic testing (GAST) may not be practical in many clinical situations. Baseline tests of ovarian reserve are easier to obtain in a routine clinical situation. Hence, we also performed logistic regression modelling excluding the day 3 (stimulated) parameters. Using multivariate logistic regression and applying stepwise forward selection on all variables summarized in Table III, we selected mean ovarian volume in the first step, and no other variable contributed significantly thereafter ($P = 0.016$) (Table III). Thus, mean ovarian volume was the most significant baseline predictor of cycle cancellation.

Table IV depicts the various ovarian reserve markers studied as predictors of a poor response to simulation. The threshold used to assess each factor was the same as that determined by the ROC curve for cycle cancellation. Inhibin B (days 2 and 3 concentrations) and AMH were both predictive of cycle cancellation and a poor response to stimulation. ROC analysis was also performed to determine the discriminatory threshold for a poor response for inhibin B and AMH (rather than using the level determined for cycle cancellation). Those thresholds are summarized in Table IV.

Although the ratio of E_2 rise (GAST) and the AFC were not predictive of cycle cancellation, they were predictive of a poor response. Just as baseline FSH concentrations were not predictive of cycle cancellation, neither were they predictive of a poor response.

Univariate logistic regression analysis showed that several variables were predictive of poor response (Table V). Age and baseline FSH levels were again not well correlated with

Table III. Logistic regression analysis for the prediction of IVF cycle cancellation expressed as odds ratios for age and study variables

Variable	Area under the receiver operating characteristic curve	Odds ratio [95% confidence interval (95% CI)]	*P*
Univariate analysis			
Age	0.54	NA	NA
Baseline FSH level (IU/l)	0.64	NA	NA
GAST estradiol (E_2) level day 3 (pmol/l)	0.70	1.0 (0.9–1.0)	0.03
Ratio E_2 rise days 2–3	0.60	NA	NA
ΔE_2 days 2–3 (pmol/l)	0.69	0.99 (0.99–1.0)	0.04
Inhibin B baseline (pg/ml)	0.67	0.98 (0.97–1.0)	0.04
Day 3 (pg/ml)	0.78	0.98 (0.97–0.99)	0.01
Sum (pg/ml)	0.77	0.99 (0.98–1.0)	0.01
Anti-Müllerian hormone (ng/ml)	0.78	0.16 (0.04–0.61)	0.01
Antral follicle count	0.74	0.76 (0.61–0.95)	0.02
Mean ovarian volume (ml)	0.78	0.71 (0.53–0.94)	0.02
Multivariate analysis			
Inhibin B day 3 (pmol/l)		0.98 (0.96–0.99)	0.017
Multivariate analysis baseline variables only			
Mean ovarian volume (ml)		0.71 (0.53–0.94)	0.016

GAST, test of ovarian response to GnRH agonist (GnRHa) administration; NA, not applicable.

Table IV. Receiver operating characteristic (ROC) curve analysis of study variables for the prediction of poor response using the same threshold as determined by the ROC area under the curve (AUC) for cycle cancellation

Variable	Threshold	Sensitivity	Specificity	Likelihood ratio (LR) [95% confidence interval (95% CI)]	*P*	New threshold
Baseline FSH level (IU/l)	≥10	69	17	1.9 (0.8–4.4)	0.13	
GAST estradiol (E_2) level day 3 (pmol/l)	≤355	55	66	1.6 (1.0–2.7)	0.17	
Ratio E_2 rise days 2–3	≤2.0	49	83	2.9 (1.3–6.3)	0.01	
ΔE_2 days 2–3 (pmol/l)	≤195	55	66	1.6 (1.0–2.7)	0.17	
Inhibin B baseline (pg/ml)	≤30	50	78	2.3 (1.1–4.4)	0.009	60
Day 3 (pg/ml)	≤60	42	90	4.2 (1.4–13.0)	0.004	87
Sum (pg/ml)	≤65	39	90	3.9 (1.2–12.0)	0.01	133
Anti-Müllerian hormone (ng/ml)	≤1.25	58	75	2.3 (1.3–4.4)	0.002	1.48
Antral follicle count	≤5	34	89	3.1 (1.1–8.4)	0.04	
Mean ovarian volume (ml)	≤5.3	41	79	2.0 (0.9–4.3)	0.06	

GAST, test of ovarian response to GnRH agonist (GnRHa) administration.

Table V. Logistic regression analysis for the prediction of IVF poor response, expressed as odds ratios

Variable	Odds ratio [95% confidence interval (95% CI)]	*P*
Univariate		
Age (years)	NA	NA
Baseline FSH (IU/l)	NA	NA
GAST estradiol (E_2) level day 3 (pmol/l)	1.00 (0.99–1.00)	0.006
Ratio E_2 rise days 2–3	0.47 (0.26–0.83)	0.009
ΔE_2 days 2–3 (pmol/l)	1.00 (0.99–1.00)	0.003
Inhibin B baseline (pg/ml)	0.99 (0.98–1.00)	0.007
Day 3 (pg/ml)	0.99 (0.98–1.00)	0.001
Sum (pg/ml)	0.99 (0.99–1.00)	0.003
Anti-Müllerian hormone (AMH) (ng/ml)	0.35 (0.19– 0.64)	0.001
Antral follicle count	0.74 (0.62–0.88)	0.001
Mean ovarian volume (ml)	NA	NA
Multivariate		
AMH (ng/ml)	0.36 (0.18–0.70)	0.003

GAST, test of ovarian response to GnRH agonist (GnRHa) administration; NA, not applicable.

response to rFSH stimulation. Neither was mean ovarian volume predictive of a poor response. Using multivariate logistic regression and applying stepwise forward selection on all variables summarized in Table V, we selected AMH in the first step, and no other variable contributed significantly thereafter ($P = 0.003$).

Pregnancy and cycle outcomes

LRs were calculated for the prediction of pregnancy using the same thresholds for each variable that were determined to predict cycle cancellation. None of the markers of ovarian reserve were able to predict pregnancy ($P > 0.05$). This suggests that these markers measure oocyte quantity rather than directly identifying deterioration in oocyte quality.

Thirteen of the 84 patients (15%) had their cycle cancelled because of a poor response. Of these, 11 patients were suitable and opted for IUI treatment. One of these patients achieved an ongoing pregnancy (9%) (Table I). Sixty-eight patients had an embryo transfer (81%). The ongoing pregnancy rate per cycle started was 11% (9/84).

Any patient who had three or less mature follicles (>14 mm) with a lead follicle at 18 mm was advised that proceeding with oocyte recovery would entail a risk of there being no embryo for transfer. Sixteen patients opted to proceed with oocyte retrieval in this situation, and 13 of 16 (81%) had an embryo for transfer and 6 of 16 (37.5%) conceived. However, 3 of 16 (19%) did not have an embryo for transfer after proceeding in this situation. The small numbers involved render it difficult to make any recommendations regarding optimizing management in this group.

Inter-cycle variability

Twenty-three patients had measures taken in a subsequent cycle to compare inter-cycle variation in the endocrine and ultrasound parameters measured in this study. All of the second cycles took place within 1 year of the original cycle. The only parameter that showed significant variation between cycles was the baseline inhibin B level (Table VI).

Discussion

In this study, we identified a group of patients at high risk of a poor response and cycle cancellation, according to their age, early follicular phase FSH concentration and a history of poor response to superovulation. Fifty-seven per cent of patients went on to have a poor response to stimulation, and 15% of stimulation cycles were cancelled. Only eight patients (10%) had eight or more oocytes collected, which is our clinic's overall mean.

Several of the markers studied were helpful in predicting the likelihood of cycle cancellation. Low serum inhibin B

Table VI. Inter-cycle variability of predictors

Variable	Cycle 1 (mean ± SD)	Cycle 2 (mean ± SD)	*P*
Baseline estradiol level (E_2) (pmol/l)	182 ± 87	145 ± 61	0.06
E_2 level day 3 (pmol/l)	430 ± 61	370 ± 203	0.11
Baseline FSH level (IU/l)	8.9 ± 3.5	9.6 ± 4.0	0.24
Inhibin B baseline (pg/ml)	67 ± 56	42 ± 45	0.04
Inhibin B day 3 (pg/ml)	173 ± 139	112 ± 83	0.08
Anti-Müllerian hormone (ng/ml)	1.36 ± 0.78	1.38 ± 0.91	0.84
Antral follicle count	7.5 ± 3.4	6.6 ± 2.5	0.43
Mean ovarian volume (ml)	4.9 ± 2.3	4.6 ± 2.4	0.64

concentrations on day 3 increased the likelihood of cycle cancellation 3-fold. Basal inhibin B concentrations would be simpler and less costly to use as a measure of ovarian reserve than stimulated day 3 concentrations. However, we observed significant inter-cycle variability in inhibin B concentration, reducing the utility of basal inhibin B measurement alone as a reliable predictor of ovarian reserve in this group of patients. Inhibin B is secreted by the developing cohort of early antral and antral follicles, the size of which varies from month to month, and thus, the amount of inhibin B being made by the granulosa cells varies. FSH concentrations also vary significantly between cycles (Scott *et al*., 1990), reflecting the relationship between circulating concentrations of inhibin B and FSH.

AMH reflects the larger resting follicular pool better than inhibin B and therefore varies less between cycles (Fanchin *et al*., 2005). There is also little variation in AMH concentration during the menstrual cycle, meaning that AMH measurement could be undertaken at any time convenient for the patient and physician (Cook *et al*., 2000). In our study, AMH was a good predictor of both cycle cancellation and a poor response to stimulation. In logistic regression analysis, AMH was the best overall predictor of a poor response to stimulation, with consistency between cycles.

AFC has recently received attention as a predictor of poor response in IVF cycles (Hendriks *et al*., 2005). In our study, AFC was confirmed as a good predictor of a poor response. However, rather surprisingly, it was not a good predictor of cycle cancellation when using ROC modelling, although AFC was predictive of cycle cancellation in the univariate analysis. It is possible that with a larger sample size, AFC would have been confirmed as predictive of cycle cancellation as the *P* value almost reached significance in the ROC model.

Ovarian volume was the best baseline predictor of cycle cancellation in logistic regression modelling when day 3 parameters were excluded. There was a trend towards ovarian volume measurement being able to predict a poor response, but this trend did not reach statistical significance. Ovarian volume did not vary significantly between cycles in this group of patients. However, there is a group of patients for whom ovarian volume measurements are impractical (e.g. those with ovarian cysts), and in our study, 9 of 84 (11%) patients were unable to have an ovarian volume measurement taken.

Dynamic testing (GAST) can be carried out at the start of a short protocol stimulation cycle. Following GAST, day 3 inhibin concentrations were predictive of cycle cancellation and poor response, as discussed above. As dynamic testing is expensive and time consuming, we suggest that these tests should not be routinely incorporated into the assessment of ovarian reserve. If dynamic testing was excluded from analysis, the best predictor of cycle cancellation was mean ovarian volume. AMH was the best predictor of a poor response in all models.

Although age and FSH levels were not significant predictors of cycle cancellation in the logistic regression model, this finding should be treated with caution. Age and previous FSH levels were enrolment criteria for the study and hence are not independent variables. The fact that this group did respond poorly confirms that they were reasonable criterion for study entry and have utility.

As predictors of cycle cancellation, only inhibin B concentration on day 3 and the mean ovarian volume had LRs >3. As predictors of a poor response, only day 3 inhibin B concentration and the AFC had an LR >3. Therefore, although several factors were statistically significant determinants of cycle cancellation and/or a poor response, the clinical utility of each test in isolation would be insufficient for them to be recommended in routine IVF practice. However, measurements may be useful in patients with a high pre-test probability of poor response and/or cycle cancellation due to reduced ovarian reserve.

The original sample size and power calculations were predicated on an odds ratio of 0.40 being achieved in the univariate logistic regression analyses. When we performed *post hoc* power calculations, only the regression models including AMH and GAST E_2 rise (the latter in the prediction of poor response only) were adequately powered (99 and 92%, respectively). Hence, some of the inconsistent results between the ROC AUC analyses and logistic regressions may have been due to an inadequate sample size. In other words, some variables that were significant predictors on ROC AUC analysis were not in the logistic regression, and this may have been due to the problems with power. A larger sample size is required to achieve adequate power for most of the logistic regression analyses presented in this article.

A test or combination of tests of ovarian reserve that could be done on one visit would be optimal for both the clinician and the patient. A one-stop clinic for basic assessment of causes of subfertilty, at which a blood test and a baseline ultrasound plus a hysterosalpingogram (or other tubal patency assessment) are carried out, has been proposed (Magos *et al*., 2005). The blood test could incorporate measuring FSH, inhibin B and AMH levels. The pelvic ultrasound assessment could incorporate an AFC and/or ovarian volume measurement. A younger woman with a reduced ovarian reserve may then choose to pursue treatment sooner, and her doctor may opt for a higher dose of stimulation and/or a different stimulation protocol. Whether this information should be used to discourage women from entering IVF treatment is debatable, because ovarian reserve testing is better at predicting a reduced response to stimulation than the possibility of pregnancy (Bukman and Heineman, 2001; Abdalla and Thum, 2004). Most women will wish to pursue motherhood using their own genetic material, and advising them otherwise on the basis of ovarian reserve testing alone is inadvisable (Esposito *et al*., 2002; Abdalla and Thum, 2004). However, this information can be used as a basis for discussion of the likelihood of a poor response to stimulation and possible reduced chance of success with IVF.

Cancellation rates would have been higher, had we been stricter about who could proceed with oocyte collection. We took an autonomous approach to patient care and allowed women to proceed with an oocyte collection even if their response did not meet our criteria for HCG administration. Although ~20% of these patients did not have an embryo for transfer, 80% of them did. The patients from this group who did have an embryo transferred achieved pregnancy rates per embryo transfer (38%) equivalent to the clinic's normal rate. Even those patients who were cancelled and converted to IUI had a chance of pregnancy. Of 84 poor prognosis patients, we were able to offer treatment (IUI or IVF) to 82.

Several of the variables we studied were helpful in predicting cycle cancellation and/or a poor response to stimulation. Yet, the goal of attempting to predict cycle cancellation must be questioned, as from our study, we would not have been able to predict cycle cancellation or pregnancy with enough accuracy to stop any of our patients from pursuing treatment. Although pregnancy rates were low, they would have been considered acceptable at many clinics only 10–15 years ago. As long as patients are aware that their chances of success are reduced, it seems reasonable to offer them the chance of pregnancy. Those who do not conceive may then be able to contemplate moving to treatment with donor oocytes, adoption or stopping attempts to become parents with greater equanimity, having experienced a cycle of stimulation with poor response.

This group of patients was already known to be at high risk of having a reduced ovarian reserve and therefore a reduced response to stimulation. The cancellation rate, low oocyte yield and pregnancy rate confirmed that. None of the other tests added enough information to be a valuable addition, with all LRs <5. More research needs to be done on ovarian reserve testing in women without infertility. Almost all the evaluation of these tests has so far concentrated on the IVF subpopulation (Bukman and Heineman, 2001). Ovarian reserve testing in these women may help with treatment decisions and counselling about prognosis but should not be used to stop access to treatment. Ovarian reserve testing earlier in the infertility workup or for reproductive lifespan information may prove to be far more helpful than in the population already accessing treatment.

References

Abdalla H and Thum MY (2004) An elevated basal FSH reflects a quantitative rather than qualitative decline of the ovarian reserve. Hum Reprod 19,893–898.

Akman MA, Erden HF, Tosun SB, Bayazit N, Aksoy E and Bahceci M (2001) Comparison of agonistic flare-up-protocol and antagonistic multiple dose protocol in ovarian stimulation of poor responders: results of a prospective randomized trial. Hum Reprod 16,868–870.

Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD and te Velde ER (2002) Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. Fertil Steril 77,328–336.

Bukman A and Heineman MJ (2001) Ovarian reserve testing and the use of prognostic models in patients with subfertility. Hum Reprod Update 7,581–590.

Cheung L-P, Lam P-M, Lok IH, Chiu TT-Y, Yeung S-Y, Tjer C-C and Haines CJ (2005) GnRH antagonist versus long GnRH agonist protocol in poor responders undergoing IVF: a randomized controlled trial. Hum Reprod 20,616–621.

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

Creus M, Penarrubia J, Fabregues F, Vidal E, Carmona F, Casamitjana R, Vanrell JA and Balasch J (2000) Day 3 serum inhibin B and FSH and age as predictors of assisted reproduction treatment outcome. Hum Reprod 15,2341–2346.

Dzik A, Lambert-Messerlian G, Izzo VM, Soares JB, Pinotti JA and Seifer DB (2000) Inhibin B response to EFORT is associated with the outcome of oocyte retrieval in the subsequent in vitro fertilization cycle. Fertil Steril 74,1114–1117.

Eldar-Geva T, Robertson DM, Cahir N, Groome N, Gabbe MP, Maclachlan V and Healy DL (2000) Relationship between serum inhibin A and B and ovarian follicle development after a daily fixed dose administration of recombinant follicle-stimulating hormone. J Clin Endocrinol Metab 85,607–613.

Esposito MA, Coutifaris C and Barnhart KT (2002) A moderately elevated day 3 FSH concentration has limited predictive value, especially in younger women. Hum Reprod 17,118–123.

Fanchin R, de Ziegler D, Olivennes F, Taieb J, Dzik A and Frydman R (1994) Exogenous follicle stimulating hormone ovarian reserve test (EFORT): a simple and reliable screening test for detecting 'poor responders' in in-vitro fertilization. Hum Reprod 9,1607–1611.

Fanchin R, Schonauer LM, Righini C, Guibourdenche J, Frydman R and Taieb J (2003) Serum anti-Mullerian hormone is more strongly related to ovarian follicular status than serum inhibin B, estradiol, FSH and LH on day 3. Hum Reprod 18,323–327.

Fanchin R, Taieb J, Lozano DH, Ducot B, Frydman R and Bouyer J (2005) High reproducibility of serum anti-Mullerian hormone measurements suggests a multi-staged follicular secretion and strengthens its role in the assessment of ovarian follicular status. Hum Reprod 20,923–927.

Hendriks DJ, Mol BW, Bancsi LF, Te Velde ER and Broekmans FJ (2005) 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 83,291–301.

Jenkins J, Davies D, Devonport H, Anthony F, Gadd S, Watson R and Masson G (1991) Comparison of 'poor' responders with 'good' responders using a standard buserelin/human menopausal gonadotrophin regime for in-vitro fertilization. Hum Reprod 6,918–921.

Lass A, Skull J, Mcveigh E, Margara R and Winston R (1997) Measurement of ovarian volume by transvaginal sonography before ovulation induction with human menopausal gonadotrophin for in-vitro fertilization can predict poor response. Hum Reprod 12,294–297.

Magos A, Al-Khouri A, Scott P, Taylor A, Sharma M, Buck L, Chapman L, Tsirkas P, Kailas N and Mastrogamvrakis G (2005) One stop fertility clinic. J Obstet Gynaecol 25,153–159.

Muttukrishna S, Suharjono H, Mcgarrigle H and Sathanandan M (2004) Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? Br J Obstet Gynaecol 111,1248–1253.

Padilla SL, Bayati J and Garcia JE (1990) Prognostic value of the early serum estradiol response to leuprolide acetate in in vitro fertilization. Fertil Steril 53,288–294.

Pearlstone AC, Fournet N, Gambone JC, Pang SC and Buyalos RP (1992) Ovulation induction in women age 40 and older: the importance of basal follicle-stimulating hormone level and chronological age. Fertil Steril 58,674–679.

Pellicer A, Lightman A, Diamond MP, Russel JB and Decherney AH (1987) Outcome of invitro fertilization in women with low response to ovarian stimulation. Fertil Steril 47,812–815.

Ranieri DM, Quinn F, Makhlouf A, Khadum I, Ghutmi W, Mcgarrigle H, Davies M and Serhal P (1998) Simultaneous evaluation of basal follicle-stimulating hormone and 17 beta-estradiol response to gonadotropin-releasing hormone analogue stimulation: an improved predictor of ovarian reserve. Fertil Steril 70,227–233.

Ravhon A, Lavery S, Michael S, Donaldson M, Margara R, Trew G and Winston R (2000) Dynamic assays of inhibin B and oestradiol following buserelin acetate administration as predictors of ovarian response in IVF. Hum Reprod 15,2297–2301.

Scott RT, Hofmann GE, Oehninger S and Muasher SJ (1990) Intercycle variability of day 3 follicle-stimulating hormone levels and its effect on stimulation quality in in vitro fertilization. Fertil Steril 54,297–302.

Scott R, Opsahl M, Leonardi M, Neall G, Illions E and Navot D (1995) Life table analysis of pregnancy rates in a general infertility population relative to ovarian reserve and patient age. Hum Reprod 10,1706–1710.

Seifer DB, Lambert-Messerlian G, Hogan JW, Gardiner AC, Blazar AS and Berk CA (1997) Day 3 serum inhibin-B is predictive of assisted reproductive technologies outcome. Fertil Steril 67,110–114.

Seifer DB, Scott RT Jr, Bergh PA, Abrogast LK, Friedman CI, Mack CK and Danforth DR (1999) 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 72,63–65.

Seifer DB, Maclaughlin DT, Christian BP, Feng B and Shelden RM (2002) Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 77,468–471.

Tinkanen H, Blauer M, Laippala P, Tuohimaa P and Kujansuu E (1999) Prognostic factors in controlled ovarian hyperstimulation. Fertil Steril 72,932–936.

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

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

te Velde ER and Pearson PL (2002) The variability of female reproductive ageing. Hum Reprod Update 8,141–154.

de Vet A, Laven JS, de Jong FH, Themmen AP and Fauser BC (2002) Antimullerian hormone serum levels: a putative marker for ovarian aging. Fertil Steril 77,357–362.

Vigier B, Picard JY, Tran D, Legeai L and Josso N (1984) Production of anti-Mullerian hormone: another homology between Sertoli and granulosa cells. Endocrinology 114,1315–1320.

Weissman A, Farhi J, Royburt M, Nahum H, Glezerman M and Levran D (2003) Prospective evaluation of two stimulation protocols for low responders who were undergoing in vitro fertilization-embryo transfer. Fertil Steril 79,886–892.

Winslow KL, Toner JP, Brzyski RG, Oehninger SC, Acosta AA and Muasher SJ (1991) The gonadotropin-releasing hormone agonist stimulation test—a sensitive predictor of performance in the flare-up in vitro fertilization cycle. Fertil Steril 56,711–717.

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

Submitted on April 13, 2006; resubmitted on June 26, 2006, September 7, 2006; accepted on September 11, 2006