Journal of Assisted Reproduction and Genetics, Vol. 22, No. 2, February 2005 (© 2005)
DOI: 10.1007/s10815-005-1495-3

Assisted Reproduction

Single and Repeated GnRH Agonist Stimulation Tests Compared With Basal Markers of Ovarian Reserve in the Prediction of Outcome in IVF

D. J. Hendriks,[1,4] F. J. Broekmans,[1] L. F. J. M. M. Bancsi,[1] C. W. N. Looman,[2] F. H. de Jong,[3] and E. R. te Velde[1]

Submitted December 31, 2003; accepted October 1, 2004

***Purpose*:** To study the value of a single or repeated GnRH agonist stimulation test (GAST) in predicting outcome in IVF compared to basal ovarian reserve tests.
***Methods*:** A total of 57 women was included. In a cycle prior to the IVF treatment, on day 3, an antral follicle count (AFC) was performed and blood taken for basal FSH, inhibin B and E_2 measurements, followed by a subcutaneous injection of 100 μg triptorelin for the purpose of the GAST. Twenty-four hours later blood sampling was repeated. All the tests were repeated in a subsequent cycle. From the GAST E_2 and inhibin B response were used as test parameters. The outcome measures were poor ovarian response and ongoing pregnancy. Group comparisons were done using the Mann–Whitney or chi-square test. Univariate and multivariate logistic regression was applied to assess which test revealed the highest predictive accuracy as expressed in the area under receiver-operating characteristic curve (ROC_{AUC}). Clinical value was compared by calculating classical test characteristics for the best logistic models.
***Results*:** All the basal and GAST variables were significantly different in the poor responders ($n = 19$) compared to normal responders ($n = 38$). In the univariate analysis on cycle 1 tests the AFC was the best predictor for poor ovarian response, while in cycle 2 the E_2 response in the GAST performed best (ROC_{AUC} of 0.91 for both). Multivariate analysis of the basal variables led to the selection of AFC and inhibin B in cycle 1, yielding a ROC_{AUC} of 0.96. Mean E_2 response was selected in a multivariate analysis of the repeated GAST variables (ROC_{AUC} 0.91). At a specificity level of ~0.90, several logistic models including GAST variables appeared to have a sensitivity (~0.80), positive predictive value (~0.82) and false positive rate (~0.18), comparable to a logistic model containing AFC and inhibin B. None of the test variables showed a significant relation with ongoing pregnancy.
***Conclusions*:** The GAST has a rather good ability to predict poor response in IVF. However, comparing the predictive accuracy and clinical value of the GAST with a day 3 AFC and inhibin B, it appeared that neither a single nor a repeated GAST performed better. In addition, the predictive ability towards ongoing pregnancy is poor. Therefore, the use of the GAST as a predictor of outcome in IVF should not be advocated.

KEY WORDS: Antral follicle count; GnRH agonist stimulation test; inhibin B; in vitro fertilization; ovarian reserve.

[1] Department of Reproductive Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands.
[2] Department of Public Health, Faculty of Medicine and Health Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands.
[3] Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands.
[4] To whom correspondence should be addressed; e-mail: d.hendriks@azu.nl.

INTRODUCTION

The current trend towards delayed childbearing has decreased the reproductive potential of women attempting to conceive. Ageing of the ovary seems to play the key role in the age related decline in ovarian reserve, which consists of depletion of the primordial follicle number, oocyte quality deterioration, decrease in implantation rates, increased risk of embryonic chromosomal abnormalities and subsequent miscarriage (1). Evaluation of ovarian reserve has become important as a screening test before entering arduous and expensive programmes of infertility therapy.

Because chronological age is of limited value in predicting individual IVF outcome, both in terms of ovarian follicular response and pregnancy rates, other markers of ovarian reserve have been studied. Basal FSH was the first widely used endocrine marker of ovarian reserve (2). As a predictor of ovarian response in IVF basal FSH generally demonstrates a modest performance. With regard to the occurrence of pregnancy its predictive ability is rather limited (3,4).

Several other endocrine ovarian reserve tests were proposed in the past decade. Clomiphene citrate challenge test (CCCT) (5–7), gonadotropin-releasing hormone agonist stimulation tests (GAST) (8), exogenous follicle stimulating hormone ovarian reserve test (EFORT) (9), basal inhibin B (10) and recently anti-Müllerian hormone (AMH) (11,12) have all been studied in order to assess their predictive value towards response to gonadotropins and chances of conception in IVF. Also from the field of quantitative ultrasonography achievements in the development of tests that assess ovarian reserve have been reported on (13,14). Moreover, the use of the combination of the basal ovarian reserve tests FSH, inhibin B and antral follicle count (AFC) was identified as a strong predictor of poor ovarian response in IVF (12,15).

It has been proposed that a dynamic evaluation of an endocrine response may provide a better assessment of ovarian reserve than basal hormone measurement (9,16–22). The GAST evaluates the estradiol serum concentration change from cycle day 2 to day 3 after administration of a supraphysiological dosage of a GnRH agonist, the latter causing a massive, temporary increase in pituitary secretion of FSH and LH. In response, the ovaries will increase estradiol release. The test is dependent on the pituitary production of gonadotropins and the response of the ovary to the subsequent gonadotropin stimulation and is believed to represent the size of the FSH sensitive cohort (8,21,23). So far, this dynamic test has been described to have good predictive power for IVF outcome (24). It is yet unclear whether it can be of additional predictive value to the information obtained from basal markers of ovarian response. Moreover, the value of repeated testing with the GAST has not been studied yet.

The aim of the present study was to investigate the predictive accuracy and clinical value of a single and repeated GAST for predicting ovarian response and ongoing pregnancy in IVF. Furthermore, we compared the performance of the GAST with the clinical value of currently used basal ovarian reserve markers.

MATERIALS AND METHODS

Patients

A total of 57 women underwent a GAST before controlled ovarian hyperstimulation for in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI). All participants fulfilled the following criteria: 1)] regular menstrual cycles (25–35 days); 2) presence of both ovaries; 3) no evidence of endocrine disorders (normal thyroid stimulation hormone, prolactin, testosterone and androstenedione); 4) age <46 years, and 5) written informed consent. The Institutional Review Board approved this study. Patients were classified according to the main cause of infertility (tubal, male or unexplained). In the majority of patients conventional IVF was planned ($n = 49$), whereas the remaining patients were scheduled for intracytoplasmic sperm injection (ICSI, $n = 8$).

Study Protocol

On day 3 of a spontaneous cycle within the 3 months preceding IVF treatment, patients underwent a transvaginal ultrasound examination to assess the number of antral follicles, measuring 2–5 mm, as described previously (15). On the same day a venous blood sample was obtained for the measurement of FSH, estradiol (E_2) and inhibin B. Serum and plasma samples were centrifuged and stored at -20°C until assayed.

A GAST was performed (21), using 0.1 mg of triptorelin s.c. (Decapeptyl; Ferring, Hoofddorp, The Netherlands). Blood samples were taken for measurement of E_2 and inhibin B immediately before

and 24 h after GnRH agonist administration. After one wash-out cycle, a second GAST was performed. The administered dose is assumed to provide maximal stimulation of the pituitary with mean peak levels of 52 IU/L for LH and 25 IU/L for FSH at 4 h after the injection and LH levels of 12.0 IU/L and FSH levels of 10.3 IU/L after 24 h (25). In this study we report on the baseline data (before stimulation) and on the results after a single and repeated GAST (delta E_2 and delta inhibin B).

Hormone Assays

E_2 and FSH were assessed in plasma with the AxSYM immunoanalyser (Abbot Laboratories, Abbot Park, IL). The E_2 assay is standardized to gas chromatography/mass spectrometry. Between run variation of E_2 assay at 300, 1105 and 2626 pmol/L was 12.5%, 7.5% and 4.9% respectively ($n = 29$). The standard of the FSH assay is calibrated against the WHO Second International Reference Preparation for human FSH (78/549). For FSH the interassay variation was 6.0, 6.6, and 8.0% at levels of 5.0, 25, and 75 IU/L ($n = 46$). Serum inhibin B levels were determined using an immunoenzymometric assay (Serotec, Oxford, UK) as described by Groome *et al.* (26) Intra- and interassay coefficients of variation were <14.6 and <14.0% respectively. For the purpose of the GAST all samples were analyzed in the same run.

Treatment Protocol

The IVF treatment has been described in detail in a previous publication (27). The patients were treated with the long downregulation protocol consisting of leuprolide acetate (Lucrin; Abbott, Hoofddorp, The Netherlands) s.c., starting in the midluteal phase of the ovarian cycle. After menstruation, a fixed daily dose of 150 IU follitropin alpha (Gonal-F; Serono Benelux BV, The Hague, The Netherlands) was administered. After 7 days follicular growth was assessed by ultrasound and E_2 measurement. If necessary the dose of rFSH was adjusted. When at least three leading follicles developed, human chorionic gonadotropin (hCG) (Profasi; Serono Benelux BV) was given to induce ovulation. Oocytes were retrieved transvaginally 36 h later. A maximum of two embryos was replaced in women <38 years of age and three above that age. The luteal phase was supported either by hCG (Profasi; Serono Benelux BV) or micronized progesterone (Progestan; Nourypharma BV, Oss, The Netherlands).

Outcome Measures

The main outcome measure of the study was poor ovarian response. As described previously (15), poor ovarian response was defined as fewer than four oocytes at follicle puncture or as cancellation due to impaired (fewer than three follicles) or absent follicular growth in response to ovarian hyperstimulation. We adopted this definition because, at a mean fertilization rate of 50–60% in IVF, a minimum of four oocytes is necessary to transfer at least two embryos, which was the intended number to be transferred in most women (15). The same definition was used in many other studies (28,29) and seems to be the most widely used definition of poor response. Moreover, defining poor responders at a slightly higher or lower cutoff yield comparable proportions of poor responders and will produce similar predictive values (30). In this study, in the group of "normal" responders we also included patients with cancelled cycles due to an exaggerated response.

A secondary outcome measure was ongoing pregnancy, defined as a viable pregnancy assessed by ultrasound of at least 11 weeks gestation. Data from patients of whom the cycle was cancelled due to either risk of ovarian hyperstimulation syndrome (OHSS) or poor response (fewer than three follicles) to hormone stimulation were not included in the pregnancy analysis, because it cannot be excluded that such patients would have become pregnant if IVF were performed. However, patients with complete absence of follicle growth and E_2 <200 pmol/L were considered to have a zero chance of pregnancy, and therefore data of their cycles were included in the analysis of pregnancy.

Methods of Analysis

Data were analyzed with the SPSS (Statistical Package for Social Sciences) for Windows (SPSS Inc., Chicago, IL). To compare normal with poor responders the Mann–Whitney test or chi-square test was performed whenever appropriate. Statistical significance was considered to be reached at $p < 0.05$. The test variables for the single GAST used in this study were delta E_2 (E_2 CD 4 minus E_2 CD 3) and delta inhibin B (inhibin B CD 4 minus inhibin B CD 3). For the repeated GAST the test variables used were mean delta E_2 (mean delta E_2 cycle 1 and

Table I. Patient Characteristics

	Total ($n = 57$)	Normal responders ($n = 38$)	Poor responders ($n = 19$)	p-Value
Age (years)	33.8 (22.3–44.0)	33.3 (26.4–44.0)	37.1 (22.3–43.3)	0.07[a]
Duration of infertility disorder (months)	30.0 (12.0–93.0)	30.5 (12.0–83.0)	30.0 (12.0–93.0)	0.77[a]
Diagnosis of infertility				
Tubal pathology (%)	10 (17.5)	5 (13.2)	5 (26.3)	
Male factor (%)	30 (52.6)	25 (65.8)	5 (26.3)	0.04[b]
Unexplained (%)	17 (29.8)	8 (21.1)	9 (47.4)	

Note. Values are presented as median (range) or as number (percentage).
[a] Mann–Whitney test.
[b] Chi-square test are performed to compare normal and poor responders.

cycle 2) and mean delta inhibin B (mean delta inhibin B cycle 1 and cycle 2). Results are presented as median and range unless otherwise indicated. Univariate and multivariate logistic regression with the main outcome-measure poor response and secondary outcome measure ongoing pregnancy were performed. Multivariate logistic regression analysis was applied with $p < 0.10$ for entry. For each single variable used in the univariate analysis as well as for the multivariate models, the ability to discriminate between patients with a poor response and patients with a normal response was assessed by calculating the area under the receiver operating characteristic curves (ROC_{AUC}). Values can range from 0.5 (*no predictive power*) to 1 (*perfect prediction*). The clinical value of the logistic models was analyzed and compared in terms of sensitivity, specificity, positive and negative predictive values, false positive and false negative rates.

To identify differences in the predictive capacity of the tests under study within age specific groups, the ROC_{AUC} was calculated and compared for the various tests and combination of tests in those aged under 35 and aged 35 years and over. This division was based on recent HFEA data, as published in the NICE report (31).

RESULTS

All 57 patients who underwent the GAST were eligible for analysis. Tables I and II present patient and ovarian reserve test characteristics of the complete group and of normal and poor responders separately. The 19 poor responders were somewhat older and were more often treated for the indication unexplained infertility. All basal (FSH, inhibin B and AFC) and GAST variables were statistically different from those in the 38 normal responders. In Table III outcome of treatment is depicted for normal and poor responders. There was an expected difference in the cancellation rate and number of oocytes retrieved. In one patient in the normal response group the ovum retrieval was cancelled because of the risk of OHSS. No ovum retrieval took place in 8 poor responders because of insufficient

Table II. Ovarian Reserve Test Variables

Variables	Total ($n = 57$)	Normal responders ($n = 38$)	Poor responders ($n = 19$)	p-Value[a]
Basal				
FSH cd3 (IU/L)	7.1 (3.6–58.9)	6.4 (3.6–17.1)	9.5 (5.4–58.9)	<0.001
Inhibin B cd3 (pg/mL)	103 (0–304)	114 (36–304)	57 (0–155)	<0.001
AFC cd3 (n)	8 (0–22)	11 (3–22)	3 (0–12)	<0.001
E_2 cd3 (pmol/L)	173 (50–1796)	173 (56–387)	186 (50–1796)	0.866
Single GAST				
Cycle 1 Delta E_2 (pmol/L)	217 (−706–914)	269.5 (75–914)	135 (−706–493)	<0.001
Cycle 2 Delta E_2 (pmol/L)	261 (−369–809)	370.5 (21–809)	93 (−369–297)	<0.001
Cycle 1 Delta inhibin B (pg/mL)	71 (−59–409)	102 (−59–409)	20 (−12–160)	0.002
Cycle 2 Delta inhibin B (pg/mL)	56 (−54–433)	88.5 (−6–433)	6 (−54–133)	<0.001
Repeated GAST				
Mean delta E_2 (pmol/L) (2 cycles)	264 (−311–851)	302.3 (55–851)	116.5 (−311–343)	<0.001
Mean delta inhibin B (pg/mL) (2 cycles)	66.5 (−29–421)	98.5 (−29–421)	17.5 (−7–145)	<0.001

Note. Values are presented as median (range). AFC: antral follicle count.
[a] Mann–Whitney test is performed to compare normal and poor responders.

Table III. Outcome Variables

	Total ($n = 57$)	Normal responders ($n = 38$)	Poor responders ($n = 19$)	p-value
Cancellation rate (%)	9 (15.8)	1 (2.6)	8 (42.1)	<0.001[a]
Number of oocytes ($n = 48$)[b]	7 (1–25)	9 (4–25)	2 (1–3)	NA
Implantation rate per embryo (%)	28.4 (23/81)	30.9 (21/68)	15.4 (2/13)	0.26[a]
Ongoing pregnancy rate/cycle (%)[c]	22.6 (12/53)	27.0 (10/37)	12.5 (2/16)	0.25[a]

Note. Values are presented as median (range). NA is not applicable.
[a] Mann–Whitney test is performed to compare normal and poor responders.
[b] Data for oocyte retrieval cycles (normal responders $n = 37$ and poor responders $n = 11$).
[c] Data for oocyte retrieval cycles or cycle cancellation due to complete absence of follicular response (normal responders $n = 37$ and poor responders $n = 16$).

follicle growth (0–2 follicles). The differences in implantation and ongoing pregnancy rates were obvious though not statistically different.

Univariate logistic regression analysis on cycle 1 data for the prediction of poor response revealed a good result for the basal variable AFC (Table IV). Predictive accuracy for FSH and inhibin B was somewhat lower (ROC_{AUC} 0.91 vs. 0.81 and 0.83, respectively). Age appeared to have no predictive capacity in the prediction of poor response. For both GAST variables delta E_2 and delta inhibin B the logistic regression analysis revealed a moderate accuracy for the prediction of poor response (ROC_{AUC} 0.83 and 0.75). When the performance of the various predictors in cycle 2 is analyzed the basal tests seem to perform on a somewhat lower level, while the accuracy for the GAST variables becomes better. The ROC_{AUC} for the repeated GAST variables mean delta E_2 and mean delta inhibin B (Table IV) was 0.91 and 0.81, respectively, indicating a moderate to good discriminating potential for predicting poor ovarian response.

When applying multivariate logistic regression analysis with stepwise forward selection on all the basal variables as presented in Table V, in both cycles AFC was selected in the first step followed by basal inhibin B in step two. As expected, the performance of this logistic model appeared somewhat better in cycle 1 (ROC_{AUC} 0.96 vs. 0.89, respectively), indicating a high discriminative potential towards poor response. Stepwise forward selection on the cycle 1 GAST variables resulted only in the selection of delta E_2 (ROC_{AUC} 0.83), while delta inhibin B added independent information in cycle 2, yielding a ROC_{AUC} for this model of 0.94. When applying stepwise forward selection on the two repeated GAST variables, only mean delta-E_2 was selected (ROC_{AUC} 0.91).

In the two different age groups (i.e. <35 years ($n = 30$) and ≥35 years ($n = 27$)) the ROC_{AUC} (95%CI) for a single GAST for the prediction of poor ovarian response was 0.76 (0.55–0.97) and 0.86 (0.73–1.00), respectively. The ROC_{AUC} (95%CI) for repeated GAST was 0.85 (0.68–1.00) and 0.93 (0.82–1.00) in

Table IV. Univariate Logistic Regression for Prediction of Poor Response Following Ovarian Hyperstimulation for IVF (Basal and GAST)

	Cycle 1		Cycle 2	
Variables	ROC_{AUC} (95% CI)	p-Values	ROC_{AUC} (95% CI)	p-Values
Basal				
Age (per year)	0.65 (0.49–0.81)	0.098	0.65 (0.49–0.81)	0.098
FSH cd3 (per IU/L)	0.83 (0.72–0.94)	0.004	0.66 (0.49–0.83)	0.050
Inhibin B cd3 (per pg/mL)	0.81 (0.68–0.94)	0.001	0.78 (0.65–0.92)	0.002
AFC cd3 (per follicle)	0.91 (0.83–0.99)	<0.001	0.86 (0.75–0.97)	<0.001
E_2 cd3 (per pmol/L)	0.51 (0.34–0.68)	0.282	0.57 (0.39–0.74)	0.065
Single GAST				
Delta E_2 (per pmol/L)	0.83 (0.71–0.94)	0.002	0.91 (0.84–0.99)	0.001
Delta inhibin B (per pg/mL)	0.75 (0.62–0.88)	0.011	0.87 (0.76–0.97)	0.001
	Cycle 1 + 2			
Repeated GAST				
Mean delta E_2 (per pmol/L)	0.91 (0.82–0.99)	0.001		
Mean delta inhibin B (per pg/mL)	0.81 (0.70–0.93)	0.003		

Table V. Multivariate Logistic Regression for Prediction of Poor Response Following Ovarian Hyperstimulation for IVF (Basal, Single GAST, and Repeated GAST)

Variables	Cycle 1		Cycle 2	
	ROC_{AUC} (95% CI)	*p*-Values	ROC_{AUC} (95% CI)	*p*-Values
Basal				
Step 1: AFC cd3 (per follicle) and	0.91 (0.83–0.99)	0.002	0.86 (0.75–0.97)	0.002
Step 2: inhibin B cd3 (per ng/mL)	0.96 (0.90–1.02)	0.006	0.89 (0.79–1.00)	0.056
Single GAST				
Step 1: Delta E_2 (per pmol/L) and	0.83 (0.71–0.94)	0.002	0.91 (0.84–0.99)	0.004
Step 2: Delta inhibin B (per pg/mL)	Not selected		0.94 (0.87–1.00)	0.051
	Cycle 1 + 2			
Repeated GAST				
Step 1: Mean delta E_2 (per pmol/L)	0.91 (0.82–0.99)	<0.001		

Note. $p < 0.10$ for entry.

the two age groups, respectively. The predictive capacity for the basal model (AFC + Inhibin B) was better than that for a single or repeated GAST in both age groups, as shown by a ROC_{AUC} (95%CI) of 0.91 (0.75–1.00) and 1.00 (1.00–1.00), respectively.

To compare the clinical performances of the model of basal markers with the single and repeated GAST models created by multivariate analysis, we calculated the classical test characteristics for the prediction of poor response at a cut-off probability of 0.50 for poor response (Table VI). In spite of the observed differences in the ROC_{AUC} among the models shown in the table, the clinical value of the AFC/inhibin B model in either cycle is not surpassed by models in which a single or repeated GAST is included. The models containing delta E_2 from a single or repeated GAST will not improve the sensitivity or positive predictive value at a specificity level of ~90%, while the false positive rate remains in approximately the same range.

In the analysis of the prediction of ongoing pregnancy after IVF, 53 patients could be included. None of the basal or GAST variables showed any statistically significant relationship with ongoing pregnancy (data not shown).

DISCUSSION

In the present study the E_2 response in the GAST showed a rather good predictive accuracy towards the occurrence of poor response in IVF patients. Unfortunately, some variation was observed between the two test cycles. This makes it not easy to draw conclusions for the true value of a single test. The observed fluctuation can only be attributed to chance variation, as there is no obvious biological explanation. In view of the interposition of a washout cycle and the unlikeliness that the status of the ovary changed substantially due to the first test or in the 2 months time course since the first test, it therefore is likely to assume that the true predictive value lies somewhere in between the values observed in the separate test cycles. Given this assumption the mean E_2 response from repeatedly performed GASTs will slightly improve the predictive accuracy of a single GAST. Based on the clinical value characteristics the repeated GAST will yield a modest increase in the sensitivity and positive predictive value at a given specificity level in comparison to the single GAST. Yet, to obtain this advantage the whole population needs to be tested twice. This implies that repeated

Table VI. Clinical Value of the Prediction of Poor Response for Various Logistic Test Models, at Cut-Off Level of the Probability of Poor Response of ≥0.50

Model	Abnormal tests (%)[a]	Sens	Spec	PPV	NPV	FP	FN	Correct predictions (%)	ROC_{AUC}
Basal model cycle 1 (AFC + inhibin B)	19 (33%)	0.84	0.92	0.84	0.92	3	3	51 (89%)	0.96
Basal model cycle 2 (AFC + inhibin B)	17 (30%)	0.74	0.92	0.82	0.88	3	5	49 (86%)	0.89
Cycle 1 GAST (delta E_2)	14 (25%)	0.53	0.89	0.71	0.79	4	9	44 (77%)	0.83
Cycle 2 GAST (delta E_2)	18 (32%)	0.79	0.92	0.83	0.90	3	4	50 (88%)	0.91
Repeated GAST (mean delta E_2)	19 (33%)	0.79	0.89	0.79	0.89	4	4	49 (86%)	0.91

Note. Sens: sensitivity, Spec: specificity, PPV: positive predictive value, NPV: negative predictive value, FP: false positives, FN: false negatives, ROC: area under the receiver operating curve.
[a] Number (percentage) of patients with an abnormal test result.

testing for clinical purposes is far from useful and confirms the conclusions published by Bancsi *et al.* on the value of repeatedly carrying out a measurement of basal FSH or an antral follicle count (32,33).

We also compared the performance of the GAST with basal ovarian reserve markers in the prediction of ovarian response. Most previous studies have only looked at various basal or dynamic ovarian reserve tests separately. When comparing the predictive accuracy of single basal tests with the GAST in a single cycle it appeared that the E_2 response in the GAST did not consistently improve the area under the ROC curve, especially if the antral follicle count is considered. Again, the variation in the performance of the basal tests between the two test cycles forces us to assume a real predictive accuracy somewhere in the middle of the two separate test results. Using combinations of basal tests, especially the antral follicle count and basal inhibin B, a level of accuracy is obtained that cannot be surpassed by measuring the E_2 response in a GAST, even if this latter test is performed twice. Moreover, the clinical value in terms of sensitivity and specificity will not clearly change by the use of the GAST. As yet, it can be concluded that the accuracy and clinical value of the GAST in the prediction of outcome in IVF does not reach a level of supremacy in comparison to basal tests that justifies the increased burden put upon the patient by subjecting them to a GAST.

There are only few studies that compared the GAST with basal ovarian reserve tests (20–22,34,35). Ranieri *et al.* compared the relative accuracy of the GAST with basal FSH and basal E_2 to predict ovarian response. They concluded that simultaneous evaluation of the basal level of FSH and the pattern of the E_2 response in the GAST can be used as a marker of diminished ovarian reserve and as a sensitive predictor of response to ovarian stimulation in patients undergoing IVF treatment. From the ROC analysis it can however be questioned whether the additional value of basal FSH is relevant in view of the good performance of the E_2 response in the GAST (21). Galtier-Dereure *et al.* (20) compared stimulated FSH in the GAST with basal FSH concentrations to predict ovarian response. They concluded that this buserelin test is strongly predictive of stimulation outcome, but is not more informative than the usual screening with basal FSH. Ravhon *et al.* (22) studied the value of basal and dynamic tests of inhibin B and estradiol in predicting ovarian response to stimulation. They showed that dynamic measurements of inhibin B and estradiol following a single administration of buserelin acetate correlated better with ovarian response to stimulation than basal FSH, inhibin B and age. Gülekli *et al.* (34) compared the accuracy of basal FSH, clomiphene stimulated FSH and the estradiol response in the GAST for predicting the number of small antral follicles within the ovaries as assessed by histology, in women who underwent oophorectomy. They concluded that none of the tests accurately reflected ovarian reserve. Finally, in the study of Ozkaya *et al.* the CCCT, GAST, basal FSH and ovarian volume were compared in a randomized fashion. This study, however, was done in an ovulation induction population treated with low dose recombinant FSH, with endpoint number of mature ($\geq$14 mm) follicles and number of recombinant FSH ampoules required for successful ovulation induction. Although the authors concluded that the ultrasound variable provided the best prediction of the two endpoints, this study seems not comparable to those carried out in IVF patients, where ovarian response is expressed as the number of follicles or oocytes, observed or obtained, after maximal ovarian stimulation with exogenous FSH (35).

To our knowledge the comparison of the GAST with AFC in the present study, is the first report in a patient group to indicate that the number of antral follicles assessed by transvaginal ultrasound delivers the same information as provided by an endocrinological challenge of this cohort. In volunteer groups a high correlation between the AFC and the estradiol response to exogenous GnRH agonist has been reported (36).

Ovarian reserve tests are most commonly used as a screening test for the presence of diminished ovarian reserve. For IVF patients this implies that the test is used for prediction of the occurrence of poor response in a standard stimulation regime. Such a test should be simple, safe, cheap and not too demanding for the patient. We consider stimulation tests generally to be unfit for this purpose. Also repeating of the test may cause delay of initiation of treatment. The combination of basal tests provides the possibility to screen larger number of IVF patients with an adequate performance in the prediction of poor responders in the subsequent stimulation. Especially the combination of antral follicle count, basal inhibin B and basal FSH has been found to be highly predictive of ovarian response in IVF treatment (12,15).

From the results in this study it can be concluded that the inhibin B response in the GAST will not consistently add relevant information to the E_2 response, which can be considered as the classical test

property. This finding seems true for both the single and repeated test variables. There are only two other studies that looked at inhibin B in the GAST (22,36). Scheffer *et al.* (36) found that in normal fertile women both the inhibin B and the E_2 responses to a single GnRH-agonist administration are most strongly correlated with the number of antral follicles. As the antral follicle count itself showed the best association with chronological age in this specific group of women it was concluded that the GAST is not expected to add explicit information to the antral follicle count. The data in the present study confirm this suggestion. Ravhon *et al.* (22) also showed that the responses of estradiol and inhibin B have similar predictive properties for ovarian response to gonadotropins in IVF patients.

Analysis of GAST in age specific subgroups has not been carried out in the literature known to date. As has been shown in the application of the CCCT in IVF cases, test value may increase if applied in specific subgroups (37,38). In the present study it was shown that overall predictive accuracy may increase in older patients. Logistic regression did not identify age as an independent predictor of poor ovarian response in addition to the basal markers or the GAST. Taken together with the limited number of cases this should caution against drawing distinct conclusions. Moreover, it was shown that the GAST either carried out once or repeated will not improve the value of basal testing within an age specific subgroup.

Like in other studies, application of the GAST (our data), AFC (15) inhibin B (10) and FSH (3,4) as a predictor of ongoing pregnancy appears to have very limited value. In the present study no differences in pregnancy and implantation rates were found between normal and poor responders, but this could be due to the small sample of patients studied. Ovarian reserve tests generally represent the quantitative aspect of ovarian reserve, whereas pregnancy is much related on the oocyte quality and will also depend on other, mostly ill-known, factors. Nevertheless, the ability to predict poor response may be a valuable tool for patient counseling, since poor responders have a clearly lower probability of pregnancy, especially if this poor response is repeatedly observed after exposure to high dosages of exogenous FSH (5,39,40). Therefore, the prediction of poor response may lead to adaptation of the stimulation dose in the first cycle to ensure that if a poor response is observed this can really be attributed to diminished ovarian reserve in stead of underdosing. In the former case continuation of IVF treatment should be discouraged strongly.

In conclusion, in this study it was shown that the clinical value of the GAST for ovarian response prediction in IVF is slightly improved by repeated testing. However, when comparing both the single and repeated GAST with basal ovarian reserve tests like the antral follicle count it appears not justified to promote the GAST as a tool for the prediction of poor ovarian response in IVF.

REFERENCES

1. te Velde ER, Pearson PL: The variability of female reproductive ageing. Hum Reprod Update 2002;8:141–154
2. 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
3. Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, te Velde ER: Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil Steril 2000;73:552–557
4. Bancsi LF, Broekmans FJ, Mol BW, Habbema JD, te Velde ER: Performance of basal follicle-stimulating hormone in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. Fertil Steril 2003;79:1091–1100
5. Loumaye E, Billion JM, Mine JM, Psalti I, Pensis M, Thomas K: Prediction of individual response to controlled ovarian hyperstimulation by means of a clomiphene citrate challenge test. Fertil Steril 1990;53:295–301
6. Navot D, Rosenwaks Z, Margalioth EJ: Prognostic assessment of female fecundity. Lancet 1987;2:645–647
7. Scott RT, Leonardi MR, Hofmann GE, Illions EH, Neal GS, Navot D: A prospective evaluation of clomiphene citrate challenge test screening of the general infertility population. Obstet Gynecol 1993;82:539–544
8. Winslow KL, Toner JP, Brzyski RG, Oehninger SC, Acosta AA, Muasher SJ: The gonadotropin-releasing hormone agonist stimulation test—A sensitive predictor of performance in the flare-up in vitro fertilization cycle. Fertil Steril 1991;56:711–717
9. Fanchin R, de Ziegler D, Olivennes F, Taieb J, Dzik A, Frydman R: Exogenous follicle stimulating hormone ovarian reserve test (EFORT): A simple and reliable screening test for detecting 'poor responders' in in-vitro fertilization. Hum Reprod 1994;9:1607–1611
10. Hall JE, Welt CK, Cramer DW: Inhibin A and inhibin B reflect ovarian function in assisted reproduction but are less useful at predicting outcome. Hum Reprod 1999;14:409–415
11. Seifer DB, MacLaughlin DT, Christian BP, Feng B, Shelden RM: Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 2002;77:468–471
12. Van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH, Themmen AP: Serum anti-Mullerian

hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002;17:3065–3071
13. Chang MY, Chiang CH, Hsieh TT, Soong Yk, Hsu KH: Use of the antral follicle count to predict the outcome of assisted reproductive technologies. Fertil Steril 1998;69:505–510
14. Tomas C, Nuojua-Huttunen S, Martikainen H: Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. Hum Reprod 1997;12:220–223
15. Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD, te Velde ER: Predictors of poor ovarian response in in vitro fertilization: A prospective study comparing basal markers of ovarian reserve. Fertil Steril 2002;77:328–336
16. Avrech OM, Royburt M, Sabah G, Zukerman Z, Pinkas H, Amit S, Ovadia J, Fisch B: The initial flare-up induced by gonadotropin releasing hormone agonist may serve as a predictor of ovarian response in the current IVF-ET treatment cycle in normogonadotropic women aged 40–48 years. J Assist Reprod Genet 1996;13:395–400
17. Dzik A, Lambert-Messerlian G, Izzo VM, Soares JB, Pinotti JA, Seifer DB: Inhibin B response to EFORT is associated with the outcome of oocyte retrieval in the subsequent in vitro fertilization cycle. Fertil Steril 2000;74:1114–1117
18. Eldar-Geva T, Margalioth EJ, Ben Chetrit A, Gal M, Robertson DM, Healy DL, Diamant YZ, Spitz IM: Serum inhibin B levels measured early during FSH administration for IVF may be of value in predicting the number of oocytes to be retrieved in normal and low responders. Hum Reprod 2002;17:2331–2337
19. Elting MW, Kwee J, Schats R, Rekers-Mombarg LT, Schoemaker J: The rise of estradiol and inhibin B after acute stimulation with follicle-stimulating hormone predict the follicle cohort size in women with polycystic ovary syndrome, regularly menstruating women with polycystic ovaries, and regularly menstruating women with normal ovaries. J Clin Endocrinol Metab 2001;86:1589–1595
20. Galtier-Dereure F, De Bouard V, Picto MC, Vergnes C, Humeau C, Bringer J, Hedon B: Ovarian reserve test with the gonadotrophin-releasing hormone agonist buserelin: Correlation with in-vitro fertilization outcome. Hum Reprod 1996;11:1393–1398
21. Ranieri DM, Quinn F, Makhlouf A, Khadum I, Ghutmi W, McGarrigle H, Davies M, Serhal P: 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 1998;70:227–233
22. Ravhon A, Lavery S, Michael S, Donaldson M, Margara R, Trew G, Winston R: Dynamic assays of inhibin B and oestradiol following buserelin acetate administration as predictors of ovarian response in IVF. Hum Reprod 2000;15:2297–2301
23. Padilla SL, Bayati J, Garcia JE: Prognostic value of the early serum estradiol response to leuprolide acetate in in vitro fertilization. Fertil Steril 1990;53:288–294
24. Bukman A, Heineman MJ: Ovarian reserve testing and the use of prognostic models in patients with subfertility. Hum Reprod Update 2001;7:581–590
25. Broekmans FJ, Bernardus RE, Broeders A, Berkhout G, Schoemaker J: Pituitary responsiveness after administration of a GnRH agonist depot formulation: Decapeptyl CR. Clin Endocrinol (Oxf) 1993;38:579–587
26. Groome NP, Illingworth PJ, O'Brien M, Pai R, Rodger FE, Mather JP, McNeilly AS: Measurement of dimeric inhibin B throughout the human menstrual cycle. J Clin Endocrinol Metab 1996;81:1401–1405
27. van Kooij RJ, Looman CW, Habbema JD, Dorland M, te Velde ER: Age-dependent decrease in embryo implantation rate after in vitro fertilization. Fertil Steril 1996;66:769–775
28. Surrey ES, Bower J, Hill DM, Ramsey J, Surrey MW: Clinical and endocrine effects of a microdose GnRH agonist flare regimen administered to poor responders who are undergoing in vitro fertilization. Fertil Steril 1998;69:419–424
29. El Toukhy T, Khalaf Y, Hart R, Taylor A, Braude P: Young age does not protect against the adverse effects of reduced ovarian reserve—An eight year study. Hum Reprod 2002;17:1519–1524
30. Van Rooij IA, Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER: Women older than 40 years of age and those with elevated follicle-stimulating hormone levels differ in poor response rate and embryo quality in in vitro fertilization. Fertil Steril 2003;79:482–488
31. Moody Y (ed): Factors affecting the outcome in in vitro fertilization treatment. *In* Fertility Assessment and Treatment for People With Fertility Problems. Clinical Guideline 11, NHS., RCOG Press, London 2004, p 85
32. Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER: Predicting poor ovarian response in IVF: use of repeat basal FSH measurement. J Reprod Med 2004;49:187–194
33. Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER: Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. Fertil Steril 2004;81:35–41
34. Gulekli B, Bulbul Y, Onvural A, Yorukoglu K, Posaci C, Demir N, Erten O: Accuracy of ovarian reserve tests. Hum Reprod 1999;14:2822–2826
35. Ozkaya O, Kaya H, Sezik M, Akyürek, Ozbasar D: The value of laboratory tests and ultrasonography in evaluating ovarian response to ovulation induction treatment with low-dose recombinant follicle-stimulating hormone. Int J Fertil Womens Med 2004;49:83–87
36. Scheffer GJ, Broekmans FJ, Looman CW, Blankenstein M, Fauser BC, de Jong FH, te Velde ER: The number of antral follicles in normal women with proven fertility is the best reflection of reproductive age. Hum Reprod 2003;18:700–706
37. Csemiczky G, Harlin J, Fried G: Predictive power of clomiphene citrate challenge test for failure of in vitro fertilization treatment. Acta Obstet Gynecol Scand 2002;81:954–961
38. Tanbo T, Dale PO, Lunde O, Norman N, Abyholm T: Prediction of response to controlled ovarian hyperstimulation: A comparison of basal and clomiphene citrate-stimulated follicle-stimulating hormone levels. Fertil Steril 1992;57:819–824
39. Lawson R, El Toukhy T, Kassab A, Taylor A, Braude P, Parsons J, Seed P: Poor response to ovulation induction is a stronger predictor of early menopause than elevated basal FSH: a life table analysis. Hum Reprod 2003;18:527–533
40. Roest J, van Heusden AM, Mous H, Zeilmaker GH, Verhoeff A: The ovarian response as a predictor for successful in vitro fertilization treatment after the age of 40 years. FertilSteril1996;66:969–973