Human Reproduction Vol.20, No.1 pp. 163–169, 2005 doi:10.1093/humrep/deh553
Advance Access publication October 7, 2004

Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve*

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

[1]Department of Reproductive Medicine, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, Departments of [2]Internal Medicine and [3]Public Health, Erasmus Medical Center, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

[4]To whom correspondence should be addressed. E-mail: d.hendriks@azu.nl

BACKGROUND: The aim of this study was to investigate the predictive accuracy and clinical value of performing either a single or a repeated clomiphene citrate challenge test (CCCT) in predicting poor response in IVF, compared to that of currently used basal ovarian reserve markers. METHODS: Sixty-three patients undergoing their first IVF treatment were prospectively included. After measurement of basal markers on cycle day 3 (cd3) [FSH, inhibin B and antral follicle count (AFC)], a CCCT was performed. FSH and inhibin B levels were measured on day 10 (cd10). A second CCCT was performed after a washout period of one cycle. In all patients the tests were followed by an IVF treatment. Poor response (<4 oocytes or cancellation due to impaired (<3 follicles) or absent follicular growth) was used as primary outcome measure. RESULTS: Both the single as well as the repeated CCCT markers had a rather good discriminative potential for the prediction of poor response (area under the receiver operating characteristic curve (ROC_{AUC}): FSH cd10 = 0.79, inhibin B cd10 = 0.79, mean FSH cd10 = 0.82 and mean inhibin B cd10 = 0.88). This compared well with the performance of the basal markers (FSH 0.82, inhibin B 0.72 and AFC 0.83). In a multivariate analysis on only the basal variables, FSH cd3 and AFC were selected (ROC_{AUC} 0.89). Only stepwise forward analysis on the repeated CCCT variables revealed a better discriminating potential for the prediction of poor response (ROC_{AUC} 0.92). At a specificity level of ~0.97, sensitivity and the positive predictive value were marginally improved in the CCCT models. CONCLUSIONS: Performing a CCCT (single or repeated) has a rather good ability to predict poor response in IVF. However, it appears that the predictive accuracy and clinical value of the CCCT is not clearly better than that of basal FSH in combination with an AFC. Therefore, the use of the CCCT as a predictor of outcome in IVF should not be advocated.

Key words: antral follicle count/clomiphene citrate challenge test/FSH/inhibinB/IVF/ovarian reserve

Introduction

Many tests have been developed to screen for diminished ovarian reserve. These tests are important because nowadays pregnancy is increasingly being postponed, leading to infertility due to diminished ovarian reserve. The clinical value of estimating ovarian reserve lies in the consequence that women with a very small chance of becoming pregnant should be advised not to participate in further therapeutic programmes, such as IVF (Bukman and Heineman, 2001).

It is generally known that reproductive ageing is related to both a quantitative and a qualitative reduction of the primordial follicle pool (Faddy and Gosden, 1996). The limited predictive value of age alone in estimating fecundity rates and response to the exogenous ovarian stimulation led to the search for better tests. Basal FSH (Scott *et al.*, 1989; Toner *et al.*, 1991), basal inhibin B (Hall *et al.*, 1999), ultrasonographic count of the antral follicles (AFC) (Bancsi *et al.*, 2002; van Rooij *et al.*, 2002), anti-Müllerian hormone (AMH) (van Rooij *et al.*, 2002), exogenous FSH ovarian reserve test (EFORT) (Fanchin *et al.*, 1994) and GnRH agonist stimulation test (GAST) (Winslow *et al.*, 1991) have all been performed in order to assess response to gonadotrophins and chances of ongoing pregnancy in IVF.

The best documented provocative test, the clomiphene citrate challenge test (CCCT) for assessing ovarian reserve, was first described by Navot *et al.* (1987) as a means of assessing ovarian reserve in women aged >35 years. This test evaluates the basal FSH level on cycle day 3 and the FSH response to clomiphene citrate (CC) administration from

*Presented in part as a poster presentation (P-356) at the 19th annual meeting of the ESHRE, Madrid, June 29–July 2, 2003.

cycle day 5 to 9. An exaggerated FSH response and/or an elevated basal FSH value is interpreted as a sign of diminished ovarian reserve. The test has been shown to be of value in unmasking poor responders to controlled ovarian stimulation (COS) who would not have been detected by basal FSH screening alone. Moreover, an abnormal test is associated with a reduced chance of pregnancy.

CC is able to occupy estrogen receptors and may act in an antagonistic as well as agonistic fashion (Adashi, 1984). At the hypothalamic level, CC almost certainly has an antiestrogenic effect, altering the pulsatile GnRH release resulting in an increased gonadotrophin (FSH and LH) output by the pituitary. In addition, CC may also exert a direct stimulatory effect on the pituitary and a direct inhibitory effect on the ovary (Adashi, 1984). Through negative feedback actions, exerting its effect at the hypothalamic or pituitary level, ovaries with a normal follicle reserve are assumed to be able to suppress FSH back to normal levels after having been stimulated by CC. In contrast, if FSH levels remain elevated, this is considered as an indirect sign of diminished ovarian reserve due to insufficient feedback from the ovary (Scott and Hofmann, 1995).

Several groups have evaluated the predictive value of the CCCT in IVF patients and stated that the CC-provoked response of FSH reliably predicts ovarian response and the probability of pregnancy in IVF (Tanbo *et al.*, 1989, 1990, 1992; Loumaye *et al.*, 1990; van der Stege and van der Linden, 2001; Csemiczky *et al.*, 2002; Yanushpolsky *et al.*, 2003). So far, most studies on the diagnostic and prognostic significance of the CCCT in IVF have used a selected group of patients, likely to respond poorly to gonadotrophins and with a reduced prospect for pregnancy. In the present study, all patients entering an IVF programme with a regular menstrual cycle, with both ovaries present and without evidence of endocrine disorders, were eligible for a study towards the value of the CCCT prior to IVF. A few studies have examined the intercycle variation of the CCCT. Hannoun *et al.* (1998) showed a high degree of intercycle variability. Kwee *et al.* (2004) also showed the CCCT to vary significantly from cycle to cycle and suggested that a high cycle variation implies reduced ovarian reserve. Because of this high cycle-to-cycle variability, we studied the effect of a repeated CCCT on the predictive accuracy and clinical value of this test. In analogy to the study by Hofmann *et al.* (1998) where they state that a diminished granulosa cell inhibin B production is the central physiological mechanism of the CCCT, we also looked at inhibin B in the CCCT.

The aim of this study was to investigate the predictive accuracy and clinical value of performing either a single or a repeated CCCT in predicting poor response in IVF, compared to that of currently advocated basal ovarian reserve markers.

Subjects and methods

Patients

A total of 63 patients who were going to have their first IVF treatment was prospectively included according to the following criteria: (i) regular menstrual cycles (25–35 days), (ii) presence of both ovaries, (iii) no evidence of endocrine disorders, (iv) age <46 years, and (v) 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 ($n = 53$), whereas the remaining patients were scheduled for ICSI ($n = 10$).

Study protocol

On day 3 of a spontaneous cycle (cd3), patients underwent a transvaginal ultrasound examination to assess the number of antral follicles, measuring 2–5 mm, as described previously (Scheffer *et al.*, 1999; Bancsi *et al.*, 2002; Hansen *et al.*, 2003). On the same day a venous blood sample was obtained for the measurement of FSH, E_2 and inhibin B. A CCCT (Navot *et al.*, 1987) was performed using 100 mg CC (Serophene®; Serono Benelux, The Netherlands) orally from day 5 to day 9 of the same cycle. On cycle day 10 (cd10), again blood samples were taken for measurement of FSH and inhibin B. After a washout cycle, a second transvaginal ultrasound examination and CCCT were performed. Serum and plasma samples were centrifuged at 1700 *g* within 2 h and stored at -20°C until assayed.

Hormone assays

FSH and E_2 were measured in plasma with the AxSYM immunoanalyser (Abbott Laboratories, USA). The standard of the FSH assay has been calibrated against the World Health Organization Second International Reference Preparation for human FSH (78/549). For FSH the inter-assay variation was 6.0, 6.6 and 8.0% at the levels of 5, 25 and 75 IU/l respectively ($n = 46$). The E_2 assay has been standardized to gas chromatography/mass spectrometry. Inter-assay variation of the E_2 assay at 300, 1105 and 2626 pmol/l was 12.5, 7.5 and 4.9% ($n = 29$) respectively. Serum inhibin B levels were determined using an immunoenzymometric assay (Serotec, UK) as described by Groome *et al.* (1996). Intra- and inter-assay coefficients of variation were <14.6 and <14.0% respectively.

Treatment protocol

The IVF treatment followed within 3 months of the second CCCT. A detailed description of the treatment protocol was published previously (van Kooij *et al.*, 1996). In brief, patients started with 200 μg leuprolide acetate (Lucrin®; Abbott, The Netherlands) in the mid-luteal phase to achieve pituitary desensitization. On day 2–3 of the subsequent menstruation, ovarian stimulation was started with a fixed dose of 150 IU follitropin alpha (rFSH, Gonal-F®; Serono Benelux) and was monitored by ultrasound and E_2 measurements. After 7 days the gonadotrophin dose was adjusted individually if necessary. When ≥ 3 follicles (≥ 18 mm) developed, 10 000 IU hCG (Profasi®; Serono Benelux), was administered and 36 h later oocyte retrieval was performed. in women aged <38 years ≤ 2 embryos were transferred. Above this age ≤ 3 embryos were transferred. The luteal phase was supported either by hCG (Profasi®) or micronized progesterone (Progestan®; Nourypharma BV, The Netherlands).

Outcome measures

The main outcome measure of the study was poor ovarian response. As described previously (Bancsi *et al.*, 2002; van Rooij *et al.*, 2002), poor response was defined as <4 oocytes at follicle puncture or as cancellation due to impaired (<3 follicles) or complete absence of follicular growth in response to ovarian hyperstimulation. We adopted this definition because, at a mean fertilization rate of 50–60% in IVF, ≥ 4 oocytes are necessary to obtain ≥ 2 embryos,

which was the intended number to be transferred in most women (Bancsi *et al.*, 2002). The same definition was used in many other studies (Surrey *et al.*, 1998; El Toukhy *et al.*, 2002) and seems to be the most widely used definition of poor response. Moreover, defining poor responders at a slightly higher or lower threshold yields, comparable proportions of poor responders and will produce similar predictive values (van Rooij *et al.*, 2003). In the analysis of poor ovarian response, the group of 'normal' responders could also include patients with cancelled cycles due to an exaggerated response (>30 follicles in both ovaries and/or peak E_2 $>15\,000$ pmol/l). A secondary outcome measure was ongoing pregnancy, defined as a viable pregnancy assessed by ultrasound, of ≥ 11 weeks gestation. Data from patients whose cycles were cancelled due to exaggerated or poor (<3 follicles) response were not included in the pregnancy analysis, because it cannot be excluded that such patients would have become pregnant if IVF had been performed. However, patients with complete absence of follicle growth and E_2 <200 pmol/l were considered to have a zero chance of pregnancy in that specific cycle and therefore were included in the pregnancy analysis.

Methods of analysis

Data were analysed with the Statistical Program for Social Sciences (SPSS Inc., USA) and General Linear Interactive Modelling packages (GLIM; NAG, UK). Values are presented as median and range. To compare normal with poor responders, the Mann–Whitney test or χ^2-test was performed whenever appropriate. A Wilcoxon signed rank was used to compare the CCCT variables from the first and second cycle. The test variables for the CCCT used in this study were FSH cd10 and inhibin B cd10. As basal tests for ovarian reserve female age, E_2 cd3, FSH cd3, inhibin B cd3 and AFC cd3 were used. Univariate and multivariate logistic regression analysis with the main outcome measure poor response and secondary outcome measure ongoing pregnancy were performed. For each single variable used in the univariate analysis and for the logistic 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}) (Hanley and McNeil, 1982). Values can range from 0.5 (no predictive power) to 1 (perfect prediction). $P < 0.05$ was considered statistically significant. The clinical value of the logistic models was analysed and compared in terms of sensitivity, specificity, and positive and negative predictive values.

Results

All 63 patients included were eligible for analysis. Fifty-three patients underwent an IVF treatment and 10 an ICSI treatment. No cases with previous ovarian surgery nor with presence of endometriotic ovarian cysts were included. Patient and basal ovarian reserve test characteristics of the complete group and of normal and poor responders separately are presented in Table I, as are the test characteristics derived from the CCCT. Patients with a poor response were slightly older and more often treated for unexplained infertility. In the poor responders basal FSH, inhibin B, AFC and the CCCT variables were significantly different in comparison to these tests in the normal responders. The CCCT variables from cycle 1 and cycle 2 were not significantly different (FSH cd10: $P = 0.09$; inhibin B cd10: $P = 0.45$). Treatment outcome variables in the normal and poor responder groups are shown in Table II. Whereas six out of 46 patients in the normal response group had their ovum retrieval cancelled because of the risk of ovarian hyperstimulation syndrome, no ovum retrieval took place in eight out of 17 poor responders because of insufficient follicle growth. Both ongoing pregnancy rates and implantation rates were lower in the poor response group.

Table I. Patient and ovarian reserve test characteristics in the total group of IVF patients, and in normal and poor responders separately

Variables	Total ($n = 63$)	Normal responders ($n = 46$)	Poor responders ($n = 17$)	*P*
General variables				
Age (years)	34.7 (24.4–42.6)	34.2 (24.4–42.6)	36.3 (26.5–42.5)	NS[a]
Duration of infertility disorder (months)	30.0 (12.0–240.0)	30.0 (12.0–78.0)	29.0 (12.0–240.0)	NS[a]
Diagnosis of infertility				
Tubal pathology	13 (20.6)	12 (26.1)	1 (5.9)	
Male factor	29 (46.0)	25 (54.3)	4 (23.5)	0.001[b]
Unexplained	21 (33.3)	9 (19.6)	12 (70.6)	
Basal variables				
FSH cd3 (IU/l)	6.1 (3.2–23.9)	5.6 (3.2–14.8)	11.0 (4.5–23.9)	<0.001[a]
Inhibin B cd3 (pg/ml)	100 (5–180)	115 (29–180)	75 (5–152)	0.007[a]
AFC cd3 (*n*)	8 (0–35)	11 (0–35)	4 (0–15)	<0.001[a]
Estradiol cd3 (pmol/l)	150 (1–1111)	140 (41–389)	157 (50–1111)	NS[a]
Single CCCT variables				
Cycle 1. FSH cd10 (IU/l)	6.4 (3.0–53.4)	6.1 (3.0–23.8)	12.2 (4.3–53.4)	<0.001[a]
Cycle 2. FSH cd10 (IU/l)	6.7 (3.4–35.4)	6.0 (3.4–14.0)	10.8 (4.1–35.4)	<0.001[a]
Cycle 1. Inhibin B cd10 (pg/ml)	227 (0–818)	267 (73–818)	155 (0–345)	<0.001[a]
Cycle 2. Inhibin B cd10 (pg/ml)	234 (1–920)	282 (112–920)	157 (1–359)	<0.001[a]
Repeated CCCT variables				
Mean FSH cd10 (2 cycles) (IU/l)[c]	6.5 (3.7–37.0)	6.1 (3.7–18.9)	13.5 (4.5–37.0)	<0.001[a]
Mean inhibin B cd10 (2 cycles) (pg/ml)[c]	242 (15–773)	275 (118–773)	159 (15–352)	<0.001[a]

Values are presented as median (range) or as number (percentage).
[a]Mann–Whitney test and [b]χ^2-test were performed to compare normal and poor responders.
[c]Median of mean values calculated from two cycles.
AFC = antral follicle count; NS = not significant; CCCT = clomiphene citrate challenge test; cd = cycle day.

Table II. Treatment outcome variables in normal and poor responders

Variables	Total ($n = 63$)	Normal responders ($n = 46$)	Poor responders ($n = 17$)	P
Cancellation rate (%)	14 (22.2)	6 (13.0)	8 (47.1)	0.001[a]
No. of oocytes ($n = 49$)[c]	7 (1–28)	8 (4–28)	2 (1–3)	N/A
Implantation rate per embryo (%)	21.2 (21/99)	26.0 (20/77)	4.5 (1/22)	0.03[b]
Ongoing pregnancy rate/cycle (%)[d]	27.8 (15/54)	35.0 (14/40)	7.1 (1/14)	0.05[b]
Live birth rate/cycle (%)[e]	27.8 (15/54)	35.0 (14/40)[f]	7.1 (1/14)[g]	0.05[b]

Values are presented as median (range) or as number (percentage). N/A = not applicable.
[a]χ^2-Test and [b]Fisher's exact test were performed to compare normal and poor responders.
[c]Data for oocyte retrieval cycles (overall $n = 49$, normal responders $n = 40$ and poor responders $n = 9$).
[d]Data for oocyte retrieval cycles or cycle cancellation due to complete absence of follicular response (overall $n = 54$, normal responders $n = 40$ and poor responders $n = 14$).
[e]Data for oocyte retrieval cycles or cycle cancellation due to complete absence of follicular response (overall $n = 54$, normal responders $n = 40$ and poor responders $n = 14$).
[f]Singletons ($n = 12$) and twins ($n = 2$).
[g]Singletons ($n = 1$).

Table III. Univariate logistic regression for prediction of poor response following ovarian hyperstimulation for IVF (basal and CCCT)

Variables	Odds ratio (95% CI)	P	ROC_{AUC} (95% CI)
Basal variables			
Age (per year)	1.06 (0.94–1.19)	0.38	0.56 (0.39–0.73)
Estradiol cd3 (per pmol/l)	1.00 (1.00–1.01)	0.09	0.54 (0.36–0.72)
FSH cd3 (per IU/l)	1.49 (1.20–1.86)	<0.001	0.82 (0.69–0.95)
Inhibin B cd3 (per pg/ml)	0.98 (0.96–0.99)	0.008	0.72 (0.58–0.87)
AFC cd3 (per follicle)	0.76 (0.65–0.90)	0.001	0.83 (0.73–0.94)
Mean FSH cd3 (per IU/l) (2 cycles)	1.72 (1.29–2.28)	<0.001	0.85 (0.74–0.96)
Mean inhibin B cd3 (per pg/ml) (2 cycles)	0.97 (0.95–0.99)	0.001	0.76 (0.63–0.88)
Single CCCT variables			
FSH cd10 (per IU/l)	1.19 (1.06–1.35)	0.004	0.79 (0.65–0.93)
Inhibin B cd10 (per pg/ml)	0.99 (0.98–1.00)	0.002	0.79 (0.68–0.91)
Repeated CCCT variables			
Mean FSH cd10 (per IU/l)	1.32 (1.11–1.56)	0.001	0.82 (0.68–0.96)
Mean inhibin B cd10 (per pg/ml)	0.98 (0.97–0.99)	0.001	0.88 (0.78–0.98)

CCCT = clomiphene citrate challenge test; cd = cycle day; ROC_{AUC} = area under the receiver operating characteristic curve; CI = confidence interval.

In Table III the results of the univariate logistic regression analysis for the prediction of poor response are shown. Because of the fact that for the various CCCT variables the ROC_{AUC} from cycle 1 and cycle 2 were not significantly different (FSH cd10: $P = 0.90$; inhibin B cd10: $P = 0.42$), Table III presents only the cycle 1 data. Mean inhibin B cd10 as a representative of a repeated CCCT presented the highest ROC_{AUC} of 0.88, indicating a good discriminating potential for predicting poor response. Age was not significantly related to poor response.

In a multivariate analysis (Table IV) of the basal reserve markers, FSH and AFC were selected (model 1), yielding a ROC_{AUC} of 0.89. In a multivariate analysis of the single CCCT markers, both FSH cd3 and inhibin B cd10 were

Table IV. Multivariate logistic regression for prediction of poor response following ovarian hyperstimulation for IVF (basal, single CCCT and repeated CCCT)

Models	Odds ratio (95% CI)	P	ROC_{AUC} (95% CI) (final model)
All basal variables (excluding mean FSH cd3 and mean inhibin B cd3) (model 1)			
Step 1: FSH cd3 (per IU/l) and	1.34 (1.08–1.67)	0.009	
Step 2: AFC cd3 (per follicle)	0.84 (0.71–1.00)	0.050	0.89 (0.79–0.99)
All single CCCT variables (including FSH cd3 and inhibin B cd3) (model 2)			
Step 1: FSH cd3 (per IU/l) and	1.44 (1.14–1.81)	0.002	
Step 2: Inhibin B cd10 (per pg/ml)	0.99 (0.98–1.00)	0.036	0.88 (0.78–0.99)
All repeated CCCT variables (including mean FSH cd3 and mean inhibin B cd3) (model 3)			
Step 1: Mean FSH cd3 (per IU/l) and	1.50 (1.13–2.00)	0.006	
Step 2: Mean inhibin B cd10 (per pg/ml)	0.99 (0.97–1.00)	0.020	0.92 (0.83–1.00)

CCCT = clomiphene citrate challenge test; cd = cycle day; ROC_{AUC} = area under the receiver operating characteristic curve; CI = confidence interval.

Table V. Clinical value for the prediction of poor response of various logistic models, at threshold level of the probability of poor response ≥0.50

Model	Abnormal tests[a]	Sensitvity	Specificity	PPV	NPV	FP	FN	Correct predictions[b]	ROC_{AUC}
(1) FSH cd3 + AFC cd3 (cycle 1)	13 (21)	0.65	0.96	0.85	0.88	2	6	55 (87)	0.89
(2) FSH cd3 + inhibin B cd10 (cycle 1)	13 (21)	0.71	0.98	0.92	0.90	1	5	57 (90)	0.88
(3) Mean FSH cd3 + mean inhibin B cd10 (2 cycles)	13 (21)	0.71	0.98	0.92	0.90	1	5	57 (90)	0.92

[a]Number (percentage) of patients with an abnormal test result.
[b]Values in parentheses are percentages.
PPV = positive predictive value; NPV = negative predictive value; FP = false positives; FN = false negatives; ROC_{AUC} = area under the receiver operating characteristic curve; cd = cycle day.

selected (ROC_{AUC} 0.88) (model 2). Mean FSH cd3 and mean inhibin B cd10 were selected in a multivariate analysis of the repeated CCCT markers (ROC_{AUC} 0.92) (model 3).

To compare the clinical performances of the single and repeated CCCT models and the basal model, created by multivariate analysis, we calculated the classical test characteristics for the prediction of poor response at a threshold probability of 0.50 for poor response (Table V). At a specificity level of ~0.97, the three models revealed sensitivities of 0.65, 0.71 and 0.71 respectively, with positive predictive values (PPV) of 0.85, 0.92 and 0.92 respectively.

In the analysis of ongoing pregnancy, 54 patients were included. There was a significant association between the nature of the ovarian response and the probability of achieving a pregnancy (Table II). None of the test variables, including those derived from the CCCT, showed a significant relationship with ongoing pregnancy [FSH cd10: ROC_{AUC} (95% CI) = 0.60 (0.43–0.78), $P = 0.25$; inhibin B cd10: ROC_{AUC} (95%CI) = 0.57 (0.41–0.73), $P = 0.43$]. When multivariate analysis was carried out with all the basal and CCCT variables, no combination of test variables was predictive of the outcome ongoing pregnancy.

Discussion

In this study we compared the predictive and clinical accuracy of a CCCT with those of currently used (combinations of) basal ovarian reserve markers. In the ROC analysis, the FSH level on cd10 in the single CCCT appeared to be a sensitive test in assessing ovarian reserve (ROC_{AUC} 0.79) if the ovarian response to hyperstimulation is used as outcome. However, the predictive accuracy was not clearly better than that of basal FSH (ROC_{AUC} 0.82), whereas in combined use the cd10 level did not add significant information to basal FSH. From these results it seems that existing literature on the value of the CCCT in IVF settings cannot be confirmed. The CCCT, first used by Navot *et al.* (1987), has been suggested to be more sensitive than screening with basal cd3 FSH alone in unmasking poor responders to COS. Tanbo *et al.* (1992) studied 91 women aged >35 years undergoing IVF and found an abnormal CCCT in 37 of them. The positive predictive value of an abnormal CCCT was 85% for cycle cancellation due to poor ovarian response and 100% for failing to conceive. In contrast, patients with a normal CCCT had a much lower cancellation rate (31.5%) and a much higher (11%) pregnancy rate in comparison to the patients with an abnormal CCCT. Several other studies reported similar results (Tanbo *et al.*, 1989; Loumaye *et al.*, 1990; Gulekli *et al.*, 1999; Csemiczky *et al.*, 2002; Yanushpolsky *et al.*, 2003). A possible explanation could be differences in the definition of an abnormal CCCT. In most studies the test results were considered abnormal in the case of an elevated cd3 and/or cd10 FSH. This means that basal FSH is incorporated into the test and this will make validation of cd10 FSH alone difficult. In the present study the value of cd10 FSH was analysed as to its additional value compared to the standard basal FSH. Another possible explanation for the differences in the suggested value of the CCCT in IVF could be that in most studies a selected group of patients with a high risk of decreased ovarian reserve was used (age >35 years, previous removal of one ovary or history of ovarian surgery, presence of endometriosis and previous poor response to ovarian hyperstimulation). This will lead to an overestimation of the positive predictive value and makes it difficult to extrapolate the results to a random IVF population.

In the multivariate analysis performed on all the single CCCT test variables (i.e. cd3 FSH, cd3 inhibin B, cd10 FSH, cd10 inhibin B) it appeared that in addition to basal FSH that was selected first, only cd10 inhibin B significantly improved the prediction of poor ovarian response. It has been shown that stimulated inhibin B concentrations reflect the number of FSH-sensitive follicles present at any time in the ovaries. Inhibin B, measured 2–6 days after starting COS, was highly correlated with the number of oocytes retrieved (Eldar-Geva *et al.*, 2002; Fawzy *et al.*, 2002; Yong *et al.*, 2003). Dynamic inhibin B testing with exogenous FSH (Dzik *et al.*, 2000) or a GnRH agonist (Ravhon *et al.*, 2000) appeared to correlate well with ovarian response to COS. In contrast, stimulated inhibin B levels in the CCCT may not consistently show a association with response in IVF, as inhibin B stimulation may often be submaximal and feedback on FSH output by inhibin B may be rather variable. There are only a few studies that report on inhibin B in a CCCT in IVF cases (Hofmann *et al.*, 1998; Kwee *et al.*, 2003). The study by Hofmann *et al.* (1998) correlated inhibin B concentrations with results of FSH measurements on cd10 of the CCCT. Women with diminished ovarian reserve on the basis of an abnormal CCCT (FSH cd3 and/or cd10 >10 IU/l) had lower cd3 and cd10 inhibin B levels compared to women with normal ovarian reserve. Recently, Kwee *et al.* (2003) showed that

stimulated values of E_2 and inhibin B in the CCCT had no additional value over the basal values in the prediction of the number of dominant follicles growing under maximal ovarian stimulation in IVF. As results in the literature are scarce and variable, the usefulness of measuring cd10 inhibin B as part of a CCCT should be further evaluated in larger studies.

In this study we compared the predictive and the clinical accuracy of the CCCT with the basal ovarian reserve markers, by creating logistic models for the prediction of poor ovarian response based on either single basal ovarian reserve markers, single CCCT variables or repeated CCCT variables. The results show that a repeated CCCT (with inhibin B) has a slightly better predictive accuracy towards the prediction of poor ovarian response compared to that of the model comprising basal FSH and basal AFC (ROC_{AUC} 0.92 versus 0.89). However, from a clinical point of view there is little or no advantage in performing a CCCT (either single or repeated) over the model basal FSH and AFC. In our opinion, the slight increase in positive predictive value and sensitivity does not justify the increased burden placed on the patient by performing a CCCT twice. In this comparison, differences in invasiveness of ovarian reserve tests should also be noted. The main difference between the so-called dynamic tests (e.g. CCCT) and the basal ovarian reserve tests is the fact that in dynamic testing a medical intervention is necessary, and patients need to attend the hospital twice. In the CCCT, patient compliance might also play a role since drugs need to be self-administered on 5 consecutive days. These possible drawbacks should be kept in mind when comparing ovarian reserve tests. Finally, the use of basal tests only may appear even more attractive if it becomes possible to use serum AMH as a measure of ovarian reserve, as its predictive value appears to be comparable to the AFC by transvaginal ultrasonography (van Rooij *et al.*, 2002).

Repeated application of ovarian reserve tests and its possible benefit for the predictive value has not been widely studied. Bancsi *et al.* (2004a,b) reported on the value of repeated measurements of basal FSH and AFC for the prediction of poor ovarian response in IVF. In these studies it was concluded that the impact on ovarian response predictions of neither a second basal FSH measurement nor a second AFC in a subsequent cycle is statistically significant. Also, clinical relevance appeared very limited in terms of changes in specificity and sensitivity. However, one study (Hannoun *et al.*, 1998), performed to determine the cycle-to-cycle variability of cd10 FSH in the CCCT, showed significant inter-cycle variability, but the influence of this variability on the ovarian reserve was not evaluated. Kwee *et al.* (2004) confirmed the cycle-to-cycle variability in the CCCT and showed that a high degree of variation of the CCCT was strongly correlated with a poor outcome of stimulation in IVF. However, from their study, it was not quite clear whether a repeated CCCT is of real clinical value in the prediction of outcome in IVF.

Female age alone was not shown to be of any value in the prediction of poor response based on the logistic regression analysis. This confirms the results as published by other authors (Ng *et al.*, 2000; Bancsi *et al.*, 2002; Yong *et al.*, 2003). Although the occurrence of poor response in older women is well established (Marcus and Brinsden, 1996; Ron-El *et al.*, 2000) the variability of response to gonadotrophin stimulation across women within age categories and the absence of a marked decay in response in women aged <38 years may hamper the identification of a logistic relationship between age and response. Finally, the relatively small sample size in the present study may have been an additional factor.

The ovarian reserve tests mentioned in this study and most of the tests described in the literature are not able to estimate the chance of achieving a pregnancy (Seifer *et al.*, 1997; Bancsi *et al.*, 2000, 2003; Hall *et al.*, 1999). This may not be surprising, as these tests merely represent the quantitative aspect of ovarian reserve. The occurrence of pregnancy in IVF as outcome measure is largely dependent on oocyte quality and factors such as transfer technique (van de Pas *et al.*, 2003) and endometrial factors. Oocyte quality cannot be directly assessed for obvious reasons. Many women with apparently normal ovarian reserve do not become pregnant in a given IVF cycle. Therefore, discriminating between individuals who do not get pregnant due to exhausted ovarian reserve and due to other factors is difficult. Perhaps the predictive value of the ovarian reserve tests regarding pregnancy may become significant if the chance of pregnancy in a series of IVF cycles is studied. Until such a study is carried out, the ability to predict poor response may be a valuable tool for patient counselling, since poor responders can expect a lower probability of pregnancy in consecutive cycles (Klinkert *et al.*, 2004). A further potential application for the prediction of poor response is the increase of the starting dose of gonadotrophins in predicted poor responders. However, whether increasing the starting dose will lead to an improvement of the response may become a serious matter of debate (unpublished data).

In conclusion, the results of our study show that performing a CCCT has a rather good ability to predict poor response in IVF and has an additional value above basal FSH alone, if used as a repeated test. However, when comparing the predictive and clinical accuracy of the CCCT with the model basal FSH and basal AFC, it appeared that a CCCT (either single or repeated) did not perform much better than the basal model alone. In our opinion this does not justify the increased burden placed on both patients and physicians by performing a CCCT. Therefore, the use of the CCCT should not be advocated as a predictor of outcome (both response and pregnancy) in IVF.

References

Adashi EY (1984) Clomiphene citrate: mechanism(s) and site(s) of action—a hypothesis revisited. Fertil Steril 42,331–344.

Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA and te Velde ER (2000) Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil Steril 73,552–557.

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.

Bancsi LF, Broekmans FJ, Mol BW, Habbema JD and te Velde ER (2003) 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 79,1091–1100.

Bancsi LF, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2004a) Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. Fertil Steril 81,35–41.

Bancsi LF, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2004b) Predicting poor ovarian response in IVF: use of repeat basal FSH measurement. J Reprod Med 49,187–194.

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.

Csemiczky G, Harlin J and Fried G (2002) Predictive power of clomiphene citrate challenge test for failure of in vitro fertilization treatment. Acta Obstet Gynecol Scand 81,954–961.

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, Margalioth EJ, Ben Chetrit A, Gal M, Robertson DM, Healy DL, Diamant YZ and Spitz IM (2002) 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 17,2331–2337.

El Toukhy T, Khalaf Y, Hart R, Taylor A and Braude P (2002) Young age does not protect against the adverse effects of reduced ovarian reserve—an eight year study. Hum Reprod 17,1519–1524.

Faddy MJ and Gosden RG (1996) A model conforming the decline in follicle numbers to the age of menopause in women. Hum Reprod 11,1484–1486.

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.

Fawzy M, Lambert A, Harrison RF, Knight PG, Groome N, Hennelly B and Robertson WR (2002) Day 5 inhibin B levels in a treatment cycle are predictive of IVF outcome. Hum Reprod 17,1535–1543.

Groome NP, Illingworth PJ, O'Brien M, Pai R, Rodger FE, Mather JP and McNeilly AS (1996) Measurement of dimeric inhibin B throughout the human menstrual cycle. J Clin Endocrinol Metab 81,1401–1405.

Gulekli B, Bulbul Y, Onvural A, Yorukoglu K, Posaci C, Demir N and Erten O (1999) Accuracy of ovarian reserve tests. Hum Reprod 14, 2822–2826.

Hall JE, Welt CK and Cramer DW (1999) Inhibin A and inhibin B reflect ovarian function in assisted reproduction but are less useful at predicting outcome. Hum Reprod 14,409–415.

Hanley JA and McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143,29–36.

Hannoun A, Abu MA, Awwad J, Kaspar H and Khalil A (1998) Clomiphene citrate challenge test: cycle to cycle variability of cycle day 10 follicle stimulating hormone level. Clin Exp Obstet Gynecol 25,155–156.

Hansen KR, Morris JL, Thyer AC and Soules MR (2003) Reproductive aging and variability in the ovarian antral follicle count: application in the clinical setting. Fertil Steril 80,577–583.

Hofmann GE, Danforth DR and Seifer DB (1998) Inhibin-B: the physiologic basis of the clomiphene citrate challenge test for ovarian reserve screening. Fertil Steril 69,474–477.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2004) A poor response in the first in vitro fertilization cycle is not necessarily related to a poor prognosis in subsequent cycles. Fertil Steril 81,1247–1253.

Kwee J, Elting MW, Schats R, Bezemer PD, Lambalk CB and Schoemaker J (2003) Comparison of endocrine tests with respect to their predictive value on the outcome of ovarian hyperstimulation in IVF treatment: results of a prospective randomized study. Hum Reprod 18,1422–1427.

Kwee J, Schats R, McDonnell J, Lambalk CB and Schoemaker J (2004) Intercycle variability of ovarian reserve tests: results of a prospective randomized study. Hum Reprod 19,590–595.

Loumaye E, Billion JM, Mine JM, Psalti I, Pensis M and Thomas K (1990) Prediction of individual response to controlled ovarian hyperstimulation by means of a clomiphene citrate challenge test. Fertil Steril 53,295–301.

Marcus SF and Brinsden PR (1996) In-vitro fertilization and embryo transfer in women aged 40 years and over. Hum Reprod Update 2,459–468.

Navot D, Rosenwaks Z and Margalioth EJ (1987) Prognostic assessment of female fecundity. Lancet 2,645–647.

Ng EH, Tang OS and Ho PC (2000) The significance of the number of antral follicles prior to stimulation in predicting ovarian response in an IVF programme. Hum Reprod 15,1937–1942.

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.

Ron-El R, Raziel A, Strassburger D, Schachter M, Kasterstein E and Friedler S (2000) Outcome of assisted reproductive technology in women over the age of 41. Fertil Steril 74,471–475.

Scheffer GJ, Broekmans FJ, Dorland M, Habbema JD, Looman CW and te Velde ER (1999) Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. Fertil Steril 72,845–851.

Scott RT and Hofmann GE (1995) Prognostic assessment of ovarian reserve. Fertil Steril 63,1–11.

Scott RT, Toner JP, Muasher SJ, Oehninger S, Robinson S and Rosenwaks Z (1989) Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. Fertil Steril 51,651–654.

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.

Surrey ES, Bower J, Hill DM, Ramsey J and Surrey MW (1998) Clinical and endocrine effects of a microdose GnRH agonist flare regimen administered to poor responders who are undergoing in vitro fertilization. Fertil Steril 69,419–424.

Tanbo T, Dale PO, Abyholm T and Stokke KT (1989) Follicle-stimulating hormone as a prognostic indicator in clomiphene citrate/human menopausal gonadotrophin-stimulated cycles for in-vitro fertilization. Hum Reprod 4,647–650.

Tanbo T, Abyholm T, Bjoro T and Dale PO (1990) Ovarian stimulation in previous failures from in-vitro fertilization: distinction of two groups of poor responders. Hum Reprod 5,811–815.

Tanbo T, Dale PO, Lunde O, Norman N and Abyholm T (1992) Prediction of response to controlled ovarian hyperstimulation: a comparison of basal and clomiphene citrate-stimulated follicle-stimulating hormone levels. Fertil Steril 57,819–824.

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.

van de Pas MM, Weima S, Looman CW and Broekmans FJ (2003) The use of fixed distance embryo transfer after IVF/ICSI equalizes the success rates among physicians. Hum Reprod 18,774–780.

van der Stege JG and van der Linden PJ (2001) Useful predictors of ovarian stimulation response in women undergoing in vitro fertilization. Gynecol Obstet Invest 52,43–46.

van Kooij RJ, Looman CW, Habbema JD, Dorland M and te Velde (1996) Age-dependent decrease in embryo implantation rate after in vitro fertilization. Fertil Steril 66,769–775.

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

Van Rooij IA, Bancsi LF, Broekmans FJ, Looman CW, Habbema JD and te Velde ER (2003) 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 79,482–488.

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.

Yanushpolsky EH, Hurwitz S, Tikh E and Racowsky C (2003) Predictive usefulness of cycle day 10 follicle-stimulating hormone level in a clomiphene citrate challenge test for in vitro fertilization outcome in women younger than 40 years of age. Fertil Steril 80,111–115.

Yong PY, 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 May 24, 2004; accepted on September 17, 2004