The predictive value of ovarian reserve tests for spontaneous pregnancy in subfertile ovulatory women

Haadsma, M. L.; Groen, H.; Fidler, V.; Bukman, A.; Roeloffzen, E. M. A.; Groenewoud, E. R.; Broekmans, F. J. M.; Heineman, M. J.; Hoek, Annemieke

Published in:
Human Reproduction

DOI:
10.1093/humrep/den234

Human Reproduction Vol.23, No.8 pp. 1800–1807, 2008 doi:10.1093/humrep/den234
Advance Access publication on June 21, 2008

The predictive value of ovarian reserve tests for spontaneous pregnancy in subfertile ovulatory women

M.L. Haadsma[1,6], H. Groen[2], V. Fidler[2], A. Bukman[3], E.M.A. Roeloffzen[1], E.R. Groenewoud[1], F.J.M. Broekmans[4], M.J. Heineman[1,5] and A. Hoek[1]

[1]Department of Obstetrics and Gynaecology, University Medical Center Groningen, University of Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands; [2]Department of Epidemiology, University Medical Center Groningen, University of Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands; [3]Department of Obstetrics and Gynaecology, Meander Medical Center, PO Box 1502, 3800 BM Amersfoort, The Netherlands; [4]Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands; [5]Department of Obstetrics and Gynaecology, Academic Medical Center, PO Box 22660, 1100 DD Amsterdam, The Netherlands

[6]Correspondence address. E-mail: m.l.haadsma@og.umcg.nl

BACKGROUND: The predictive value of ovarian reserve tests (ORTs) for spontaneous pregnancy is unclear. Our study aimed to determine whether ORTs have added value to previously identified prognostic factors for spontaneous pregnancy in subfertile ovulatory couples. METHODS: A prospective cohort study was performed on 474 subfertile ovulatory couples in two hospitals in Groningen, The Netherlands. The ORTs performed were: antral follicle count (AFC), follicle-stimulating hormone (FSH), inhibin B (basal levels and after stimulation with clomiphene citrate) and the clomiphene citrate challenge test. For each couple, the probability of spontaneous pregnancy was retrospectively calculated using the validated Hunault prediction model which includes the main known prognostic factors for spontaneous pregnancy. Outcome measure was time to spontaneous pregnancy resulting in a live birth. RESULTS: When added to the Hunault model, only basal FSH and AFC significantly improved the prediction of spontaneous pregnancy (*P*-values of 0.05 and 0.04). Absolute changes in predicted probabilities after adding basal FSH or AFC were small: the predicted probability of spontaneous pregnancy shifted ≥10% in only 3.8% and 7.9% of the couples, respectively. CONCLUSIONS: Although basal FSH and AFC significantly improved the validated prediction model for spontaneous pregnancy, the clinical relevance of this finding is limited. We recommend that none of the ORTs studied should be used routinely in the subfertility evaluation of ovulatory couples to predict spontaneous pregnancy chances.

Keywords: ovarian reserve test; spontaneous pregnancy; predictive value; prediction model; reproductive ageing

Introduction

In Western societies, the widespread availability of contraception and the altered social position of women have led to a rapid increase in mean age at which couples try to conceive their first child. The chance to conceive, however, decreases with female age. The trend to postpone parenthood has led to an increase in the number of couples experiencing subfertility (Te Velde and Pearson, 2002). After medical evaluation of their fertility problem, couples are either advised to pursue the possibility of spontaneous conception or start treatment. To support a well-informed decision, spontaneous pregnancy chances for couples are estimated using prediction models consisting of previously identified prognostic factors, such as female age, duration of subfertility and sperm motility (Eimers *et al*., 1994; Collins *et al*., 1995; Hunault *et al*., 2004, 2005). If a couple has a high probability of spontaneous conception, treatment does not enhance their pregnancy chances and can be regarded as an unnecessary burden for the couple. In contrast, if a couple has a low probability of spontaneous conception, delay of treatment is unfavourable, especially in cases of advanced female age. Recently, one of the prediction models for spontaneous pregnancy, referred to as the Hunault model, has been externally validated in a large population (Van der Steeg *et al*., 2007a).

The process of reproductive ageing is attributed to a decrease in quantity and quality of the ovarian follicle pool (Te Velde and Pearson, 2002). Among women of the same age, reproductive potential may vary considerably (Broekmans *et al*., 2004). To assess the individual ovarian reserve, several ovarian reserve tests (ORTs) have been developed, which provide a quantitative measure of the ovarian reserve. Most ORTs are baseline or dynamic endocrine parameters, such as the basal

level of follicle-stimulating hormone (FSH) or the clomiphene citrate challenge test (CCCT). The second group of widely used ORTs consists of sonographic tests, such as the antral follicle count (AFC). The value of ORTs has been evaluated mainly in women undergoing *in vitro* fertilization (IVF) treatment. ORTs have, for a long time, been thought to predict both ovarian response to stimulation and IVF-related pregnancy. However, recent meta-analyses in IVF-treated populations only show a moderate predictive value of ORTs for ovarian response and virtually no relation with pregnancy rate (Broekmans *et al*., 2006; Maheshwari *et al*., 2006). Less is known about the value of ORTs for the prediction of pregnancy in the general subfertile population. Although ORTs are frequently used in the basal fertility work-up of the subfertile couple, only a few studies are available on the subject, with conflicting results (Scott *et al*., 1995; Van Montfrans *et al*., 2000; Van Rooij *et al*., 2006; Van der Steeg *et al*., 2007b). The value of ORTs as part of the evaluation of the subfertile couple is thus unclear. If ORTs indeed have a predictive value for spontaneous pregnancy, it is of interest to know whether they have significant added value to previously identified prognostic factors.

In view of the ongoing debate on the value of ORTs in the management of the subfertile couple, we performed a prospective cohort study in a subfertile ovulatory population. This exploratory study was aimed to answer the question of whether ORTs are predictive for the occurrence of spontaneous pregnancy and whether they have added value to previously identified prognostic factors for spontaneous pregnancy. For this purpose, we used the validated Hunault prediction model in our analysis.

Materials and Methods

Study population

From December 1999 to July 2003, patients were recruited at the tertiary fertility centre of the University Medical Center Groningen and the fertility centre of the Martini Hospital, a teaching hospital in Groningen, the Netherlands. Patients were asked to participate after a basal subfertility evaluation had been completed. This evaluation included assessment of ovulation by sonographic cycle analysis and measurement of midluteal progesterone, semen analysis, post-coital test and hysterosalpingography. The inclusion criteria for the study participation were (i) subfertility for at least 12 months, (ii) a regular ovulatory cycle with midluteal progesterone of >30 nmol/l and a mean cycle length between 21 and 42 days, (iii) at least one patent tube at hysterosalpingography and/or laparoscopy, (iv) semen analysis with a total motile count $>1 \times 10^6$ and (v) no sexual disorder leading to a coitus frequency of less than once a month. After basal subfertility evaluation, diagnostic laparoscopy was performed if tubal pathology was suspected clinically (e.g. a history of pelvic inflammatory disease), after an abnormal result of hysterosalpingography and before starting treatment with intrauterine insemination. Laparoscopy was performed after patients were asked to participate in the study; if two-sided tubal pathology was detected, patients were excluded secondarily. Informed consent was obtained from all participants. The study protocol was reviewed and approved by the Institutional Research Boards of both participating hospitals.

Ovarian reserve tests

We chose to perform ORTs from each category, i.e. basal and dynamic endocrine and sonographic. We decided to perform basal FSH and the CCCT since the study of Scott *et al*. (1995) demonstrated a predictive value of these tests for pregnancy. We decided not to add other dynamic tests to reduce the burden for the patients and keep the protocol as short as possible, in order to increase the inclusion rates and prevent drop-out. Inhibin B was described by Hofmann *et al*. (1998) as the physiologic basis of the CCCT and was therefore added to our protocol. Since anti-Müllerian hormone was not recognized as such a useful ORT at the start of the study in 1999 as it is now, we did not perform this test and unfortunately no serum was stored. In contrast, the AFC had already been described as ORT, among others by Scheffer *et al*. (1999). In this study, the total ovarian volume showed only a moderate relation with female age and therefore seemed less suitable as an ORT compared with the AFC.

Study protocol

After inclusion, patients visited the out-patient clinic in the early follicular phase of the menstrual cycle (cycle day 2, 3 or 4). Transvaginal ultrasound was performed by one of four skilled gynaecologists using a 7.5 MHz vaginal probe on an Aloka SSD-1700 US machine. Follicles were counted and measured in two dimensions. The mean of these measurements was taken to be the follicle size. The numbers of follicles sized 2–6 mm from both ovaries were added for the total AFC (Haadsma *et al*., 2007). Peripheral blood was drawn to measure basal levels of FSH and inhibin B. Subsequently, patients self-administered 100 mg clomiphene citrate on cycle day 5–9. One week after the first visit, blood samples were drawn again to measure stimulated FSH and inhibin B levels. The result of the CCCT was defined as the sum of the values for basal and stimulated FSH.

Hormone assays

For measurements of concentrations of FSH and inhibin B, serum was stored at -20°C until processing. Serum FSH levels were measured by fluorimmunometric determination on the AutoDelfia (Wallac/Perkin Elmer, Turku, Finland). For FSH, the inter-assay coefficient of variation was 3.7%, the sensitivity <0.05 IU/l. The lower limit of detection was 0.03 IU/l. The standard of the FSH assay was calibrated against the WHO Second International Reference Preparation for human FSH (78/549). Inhibin B concentrations were assayed with Enzyme Immunoassay (ELISA) from Serotec (Kidlington, Oxford, UK). The inter-assay coefficient of variation for inhibin B was 11% and the sensitivity was <10 pg/ml. The lower limit of detection was 5 pg/ml.

Follow-up

After completion of the basal subfertility evaluation, expectant management or treatment was proposed to each couple. The advice was based on duration of subfertility and the estimated chances for spontaneous conception according to the prediction model routinely used in both clinics at the time of the study (Eimers *et al*., 1994). Couples were advised to start treatment if the estimated chance to conceive was below 30% or duration of subfertility exceeded 3 years (2 years in the case of female age $\geq$37 years). Couples that were primarily advised expectant management and did not conceive spontaneously were also offered treatment as soon as they met these criteria. The kind of treatment offered to the couples with a low chance of spontaneous conception depended on the type of subfertility. Patients were offered treatment by intrauterine insemination with stimulation in the case of unexplained subfertility and mild male factor, intrauterine insemination without stimulation in the

case of cervical factor and IVF or intracytoplasmic sperm injection in the case of severe male factor. Treatment was not generally available for women over the age of 40. For all couples, data regarding spontaneous pregnancies and start of treatment were recorded. Information was obtained from medical files and from questionnaires completed by the participating couples. Couples were excluded secondarily if they did not meet the inclusion criteria any longer. The main reason for secondary exclusion was two-sided tubal pathology on laparoscopy because of adhesions and/or severe endometriosis, despite one or two patent tubes on hysterosalpingography.

For all couples, including those who were advised to start therapy, follow-up started on the day of the first ORT. The study end-point was a spontaneous, treatment-independent pregnancy resulting in a live birth. The first day of the last menstrual cycle was recorded as the end-point of time to pregnancy. For couples who achieved a spontaneous pregnancy but miscarried, follow-up was continued. For the couples who did not achieve an ongoing spontaneous pregnancy, follow-up ended when treatment was started. Since there were hardly any waiting lists for therapy, this implies that duration of follow-up was often short for couples who were advised treatment. Follow-up also ended when couples started contraceptives or ended their relationship. Couples who never started treatment were sent a questionnaire to complete data. The last date of follow-up was 1 November 2006.

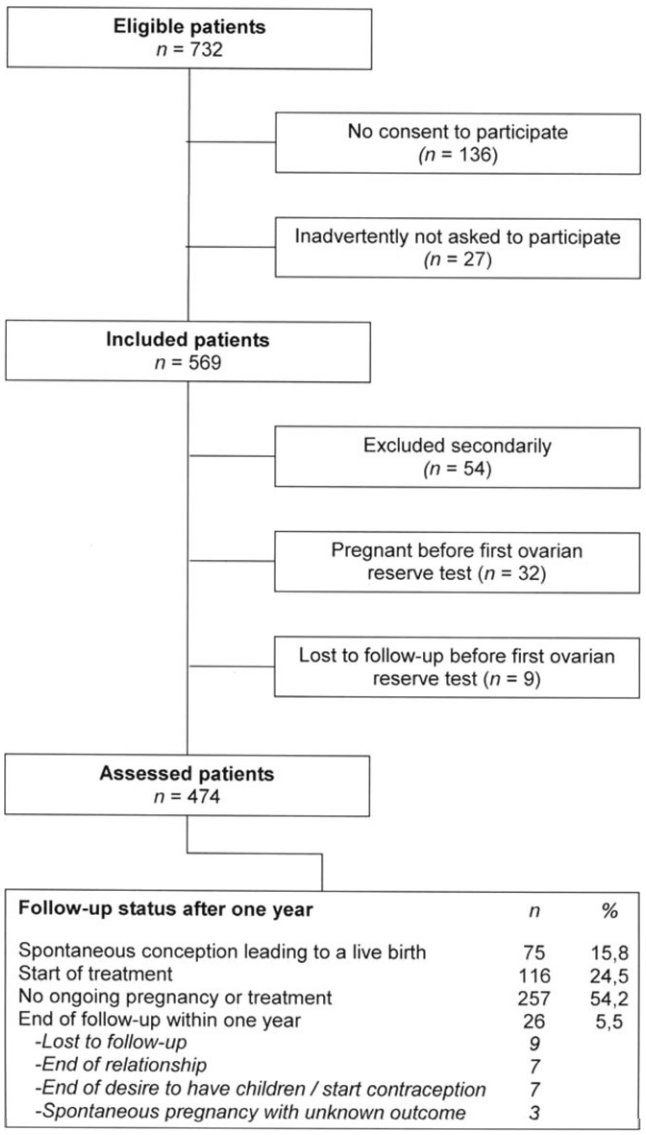
Prognostic model of Hunault

To assess the additional value of the ORTs to previously identified prognostic factors for spontaneous pregnancy, we used the prognostic model of Hunault (Hunault *et al*., 2004; Van der Steeg *et al*., 2007a). This model was developed to predict the chance of spontaneous pregnancy within 1 year leading to a live birth. It includes the factors female age, duration of subfertility, type of subfertility (primary or secondary), sperm motility and referral status (referred by a general practitioner or gynaecologist). Two versions of the model have been developed, one with the result of the post-coital test as prognostic factor and one without the post-coital test (Hunault *et al*., 2005). In our study population, the post-coital test had a significant additional value to the Hunault model for the prediction of spontaneous pregnancy (P-value = 0.026). We therefore used the version of the model including the post-coital test in our analysis except in those calculations where the value of the post-coital test itself to the model was assessed. Comparison of predicted and actual spontaneous pregnancy rates showed a reasonable predictive value of this model in our study population (c-statistic 0.67, 95% confidence interval 0.61–0.73) (Harrell *et al*., 1996).

Statistical analysis

The present study is an exploratory study; no formal power analysis was performed. To analyse the relation between the various ORTs and spontaneous pregnancy, we used Cox regression analysis. Time to pregnancy was calculated for those achieving an ongoing pregnancy resulting in a live birth. Couples who did not achieve this outcome were censored at the time that treatment was started or when follow-up ended, whichever came first. The functional relationship between each ORT and the hazard function for spontaneous pregnancy was explored by fitting smoothing splines (Therneau and Grambsch, 2000). More precisely, we fitted models which included each ORT separately as a p-spline function with five degrees of freedom. This was done with and without the Hunault score. The statistical significance of each ORT was initially assessed by likelihood ratio tests performed at a 10% level. On significance, we explored the possibility of replacing the p-spline by a simple quadratic or a piecewise linear function.

The predictive value of each ORT for the chance of spontaneous pregnancy was assessed in a univariate analysis and in a multivariate model together with the Hunault score. We calculated the individual probabilities of spontaneous pregnancy using different models in order to compare them. The calculations were performed for a model with the Hunault score only ('customized' Hunault model) and for models including the Hunault score and the ORTs that showed an additional value in the multivariate analysis. The clinical implications of the addition of an ORT to the Hunault model were assessed by calculating the number of couples for which the probability shifted 10% or more (for instance from 15% to 25% or vice versa). Second, we calculated the number of couples that would change from a probability of $\geq$30 to $<$30% or vice versa. This threshold value is used in both participating clinics: expectant management is advised to couples with an estimated chance for spontaneous conception within the coming year of at least 30%; below 30%, couples are advised to start treatment (Steures *et al*., 2006). Subsequently, the observed pregnancy rates of these couples were calculated using Kaplan–Meier analysis. Data were analysed with SPSS 14.0 (SPSS Inc., Chicago, IL, USA) and S-plus 7.0 (MathSoft Inc., Seattle, WA, USA).



Follow-up status after one year	*n*	%
Spontaneous conception leading to a live birth	75	15,8
Start of treatment	116	24,5
No ongoing pregnancy or treatment	257	54,2
End of follow-up within one year	26	5,5
-Lost to follow-up	*9*	
-End of relationship	*7*	
-End of desire to have children / start contraception	*7*	
-Spontaneous pregnancy with unknown outcome	*3*	

Figure 1: Flowchart of eligible patients.

Results

In total, 474 patients were entered in the cohort follow-up (Fig. 1); 353 in the University Medical Center Groningen and 121 in the Martini Hospital. The end-point of spontaneous conception within 1 year resulting in a live birth was achieved by 75 couples (15.8%). Using the Kaplan–Meier analysis, the spontaneous ongoing pregnancy rate after 1 year was 21.7% (standard error 2.4%). Characteristics of the study population and outcome of the ORTs are shown in Table I. Included in the study were 27 women (5.7%) who were over 40 years of age and 7 women (1.5%) who had a cycle length between 35 and 42 days. AFC was inadvertently not performed in 75 couples (15.8%); patient characteristics in the group with AFC did not differ from the group without AFC. Likewise, in 28 couples, no post-coital test was performed (5.9%).

The univariate Cox regression analysis showed that none of the ORTs had a significant predictive value for time to spontaneous pregnancy (Table II). Only basal FSH showed a borderline significant predictive value (P-value = 0.051). Subsequently, each ORT was added to the Hunault regression model. Basal FSH (P-value = 0.092), CCCT (P-value = 0.040) and AFC (P-value = 0.064) were selected for further analysis. The relations of basal FSH, CCCT and AFC with the hazard function for spontaneous pregnancy are shown in Fig. 2a–c. For stimulated FSH, basal and stimulated inhibin B, P-values were >0.10; hence they were not analysed further (Table II). As suggested by Fig. 2a, the relationship between basal FSH and the hazard function for spontaneous pregnancy was best described by a quadratic polynomial, graphically depicted as a parabola with an optimum around basal FSH = 8 IU/l. Both above and below this value, spontaneous pregnancy chances diminished. We also explored model variants that ignored the influence of the lower FSH-values, but such models did not fit to the data. The relation between AFC and the hazard function for spontaneous pregnancy (Fig. 2c) was best described by a piecewise linear regression model with one knot at AFC = 12; both above and beneath this value, spontaneous pregnancy chances declined. Further exploration ignoring the influence of higher AFC-values did not result in identification of a simplified model that was statistically significant. For the CCCT, the functional relationship with spontaneous pregnancy could not be described in a clinically meaningful manner (Fig. 2b). Simplified models assuming an optimum score of the CCCT did not add significantly to the Hunault model. Therefore, this test was not included in subsequent analyses. If AFC and basal FSH were simultaneously added to the Hunault model, the additional value of basal FSH was no longer significant, implying that AFC was a stronger independent predictor than basal FSH. The analyses were also performed using the Hunault model without the post-coital test and showed similar results.

The probabilities for spontaneous pregnancy calculated with the Hunault model and with the Hunault model including basal FSH or AFC are presented in Fig. 3a and b. The numbers of couples for which the probability would alter 10% or more after the addition of either basal FSH or AFC to the model are shown in Table III. The probabilities calculated with the Hunault model including and excluding the post-coital test are added for comparison. Also the numbers of couples for which clinical management would change (i.e. a change in probability from $\geq$30 to $<$30% or vice versa) if either basal FSH or AFC were added to the Hunault model are shown in Table III, as are their observed pregnancy rates.

Discussion

In our subfertile ovulatory population, basal FSH and AFC show a statistically significant added value to the validated Hunault prediction model, a model consisting of previously identified prognostic factors for spontaneous pregnancy. Stimulated FSH, CCCT, and basal and stimulated inhibin B have no value for the prediction of spontaneous pregnancy.

The relations of basal FSH and AFC with spontaneous pregnancy chances show a symmetric pattern with an optimum

Table I. Patient characteristics and outcome of the ovarian reserve tests.

	Median (* number)	10th–90th percentiles (* %)	*N*
Patient characteristics			
Age (years)	32.5	26.6–38.8	474
Duration of subfertility (months)	26.5	16.8–50.4	474
Duration of follow-up (days)	203	48–657	474
Primary subfertility*	324	68.4%	473
Mean cycle length (days)	28	25–32	461
Semen analysis (TMC, $\times 10^6$)	34.5	4.0–175.6	474
Diagnostic category of subfertility*			474
Unexplained	243	51.3%	
Male factor	213	44.9%	
Cervical factor	18	3.8%	
Outcome of ovarian reserve tests			
Antral follicle count (*n*)	10	4–23	399
Basal FSH (IU/l)	6.6	4.6–10.6	465
Stimulated FSH (IU/l)	6.6	4.0–11.0	466
CCCT (basal+stimulated FSH) (IU/l)	13.2	9.1–21.8	460
Basal inhibin B (ng/l)	89.0	37.7–140.0	466
Stimulated inhibin B (ng/l)	226.0	104.0–414.4	462

TMC, total motile count (volume × concentration × motility); FSH, follicle stimulating hormone; CCCT, clomiphene citrate challenge test.

Table II. Relation of ovarian reserve tests and time to spontaneous pregnancy resulting in a live birth: univariate analysis and additional value to the Hunault model.

Ovarian reserve test	Univariate analysis	Additional value to the Hunault model	
	P-value	Model form	*P*-value
Basal FSH (bFSH)	0.051	Smoothing splines	0.092
		Quadratic	0.047
Stimulated FSH (sFSH)	0.936	Smoothing splines	0.165
CCCT (bFSH + sFSH)	0.455	Smoothing splines	0.040
Antral follicle count	0.144	Smoothing splines	0.064
		Linear (node = 12)	0.037
Basal inhibin B	0.321	Smoothing splines	0.592
Stimulated inhibin B	0.147	Smoothing splines	0.393

P-values of likelihood ratio test; FSH, follicle stimulating hormone; CCCT, clomiphene citrate challenge test.

value above and below which pregnancy chances decrease (Fig. 2a and c). These relations seem to describe two clinical phenomena. One leg of the curve depicts decreased pregnancy chances for values representing diminished ovarian reserve. The other leg of the curve, however, shows decreased pregnancy chances for low values of basal FSH or high values of AFC, representing an abundance of antral follicles. Although our population was selected after ovulation detection, possibly intermittent occurrence of anovulatory cycles, such as that seen in patients with polycystic ovaries, or subclinical hormonal disturbances may cause diminished pregnancy rates in these groups. This phenomenon is at least partly responsible for the fact that basal FSH and AFC add significantly to the Hunault model, since simplified models including only the effect of high FSH or low AFC values are statistically not significant.

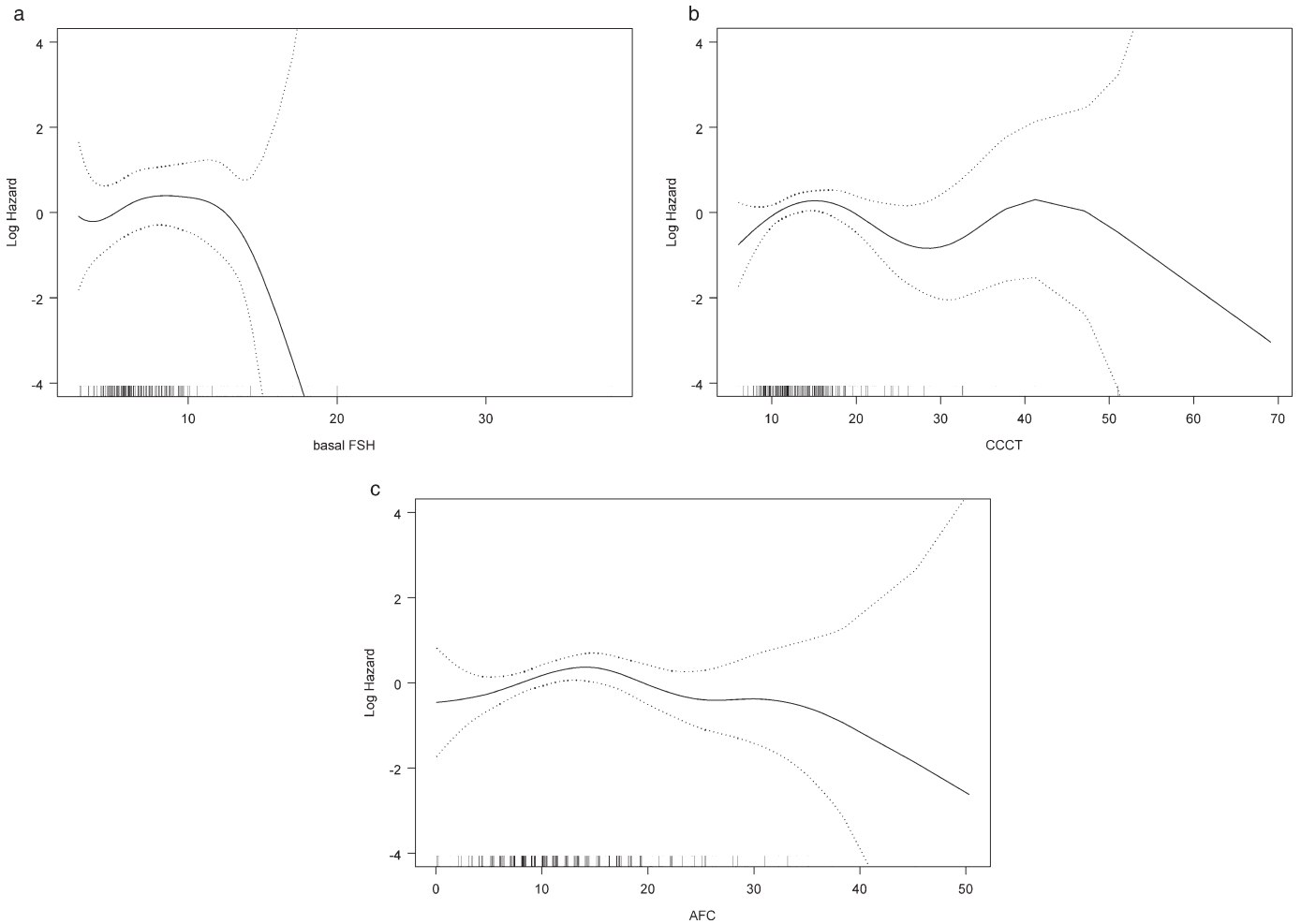


Figure 2: Graphical relation of basal FSH, CCCT and AFC with time to spontaneous pregnancy.
Relation of basal values of FSH (in IU/l) (**a**), CCCT (in IU/l) (**b**) and AFC (**c**) with the log hazard ratio for time to spontaneous pregnancy resulting in a live birth. FSH, follicle stimulating hormone; CCCT, clomiphene citrate challenge test; AFC, antral follicle count.

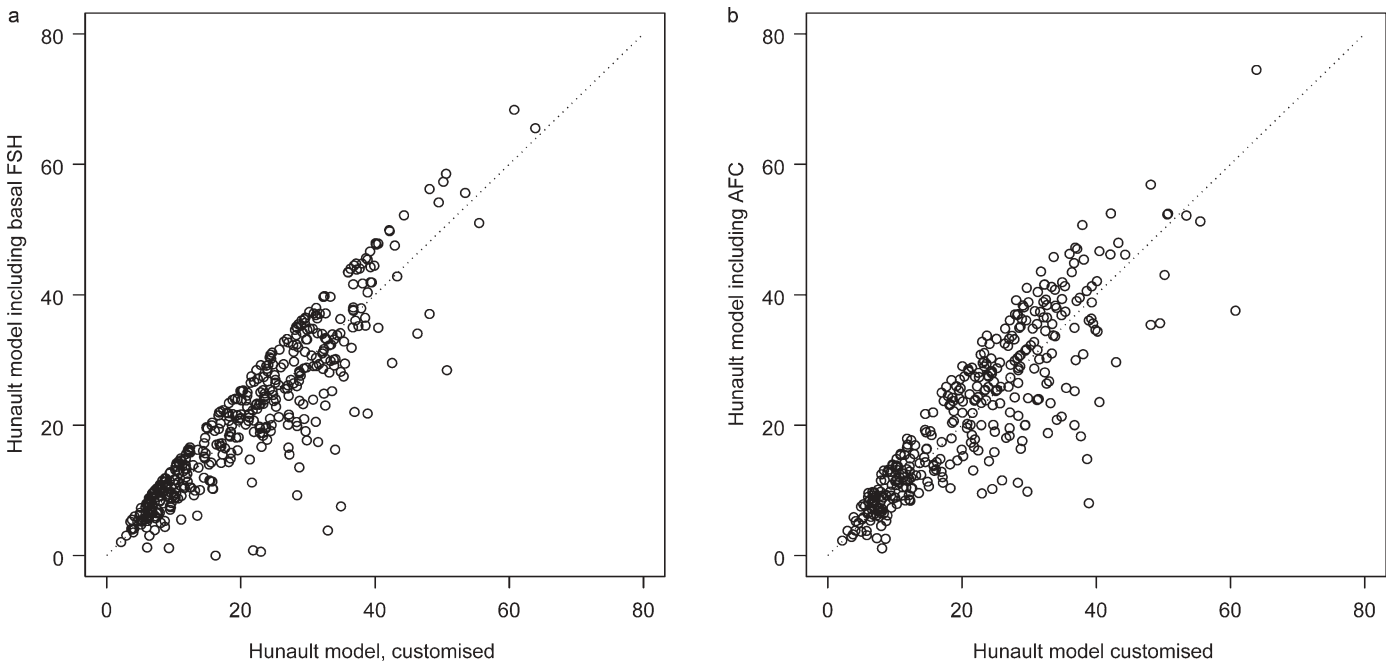


Figure 3: Relation between the predicted probabilities for spontaneous pregnancy based on the Hunault model before and after the addition of basal FSH or antral follicle count.
Probabilities (in %) for each couple of spontaneous pregnancy within 1 year resulting in a live birth based on the Hunault model and the Hunault model including basal levels of FSH (**a**) or AFC (**b**) FSH, follicle stimulating hormone; AFC, antral follicle count.

The main goal of the Hunault model is to discriminate between couples with a good and poor prognosis; the accuracy of the calculated probabilities for spontaneous pregnancy is determined by the limitations of the model. The change in probability after the addition of one of the ORTs should be interpreted in this light. If basal FSH or AFC are added to the Hunault model, the probability shifts by 10% or more for only 3.8% and 7.9% of the couples, respectively, and this shift will not be accurate in all of these couples.

Another way to assess the impact of the addition of an ORT to the model is to calculate the number of couples that would receive a different advice regarding subfertility management (i.e. a shift in probability from $<30\%$ to $\geq 30\%$ and vice versa); for basal FSH, this is 10.8% and for AFC, 13.6% of the couples. These numbers include couples with only a limited shift in probability from just below to little over 30%. The change in subfertility management is clinically relevant only if pregnancy chances are indeed better predicted for these couples, and they are more often assigned to the right treatment group. However, numbers are too small to demonstrate this, as is shown by the wide confidence intervals. More importantly, our data do not indicate that the addition of an ORT to the Hunault model identifies a substantial new group of couples with a low probability for spontaneous pregnancy that would otherwise have been unjustly advised expectant management.

We chose the post-coital test to compare the effect of the ORTs as addition to the Hunault model. The position of the post-coital test in the Hunault model is subject of debate, since the additional predictive value of this test has not been demonstrated unambiguously (Van der Steeg *et al*., 2007b). Compared with the ORTs, the effect of the addition of the

Table III. Clinical impact of the addition of basal FSH or antral follicle count to the Hunault model.

Added variable to the Hunault model	Shift in probability of 10% or more	Shift in probability from ≥ 30 to $<30\%$ New advice: start treatment		Shift in probability from <30 to $\geq 30\%$ New advice: expectant management	
	No. of couples (%)	No. of couples (%)	Observed pregnancy rate (95% CI)	No. of couples (%)	Observed pregnancy rate (95% CI)
Basal FSH	14 (3.8%)	20 (5.4%)	50% (26–74)	20 (5.4%)	23% (3–44)
Antral follicle count	29 (7.9%)	21 (5.7%)	37% (12–62)	29 (7.9%)	26% (9–42)
Post-coital test	52 (14.2%)	12 (3.3%)	0% (0–46)	35 (9.5%)	49% (30–68)

The probability of spontaneous pregnancy for each couple using the Hunault model is compared with their probability calculated with the Hunault model including basal FSH or antral follicle count. The first column shows the number of couples for which the probability of spontaneous pregnancy shifts 10% or more after addition of basal FSH or antral follicle count to the Hunault model. The second and third column show the number of couples for which the shift in probability would cause a change in advice on their subfertility management and show their actual observed pregnancy rates. The effect of the post-coital test as addition to the Hunault model is added for comparison. For this table, only those couples were taken into account in which the results of all variables of the Hunault model, basal FSH, antral follicle count and post-coital test were available ($N = 367$). CI, confidence interval; FSH, follicle stimulating hormone.

post-coital test to the Hunault model is more substantial: in 14.2% of the couples, the probabilities shift by ≥10%, compared with 3.8% and 7.9% for basal FSH and AFC, respectively.

Limitations

In our study, the results of the ORTs were known to doctors and patients. It cannot be excluded that the type of management advised or carried out was influenced by the results of the ORTs, despite the clear criteria as described in the Materials and Methods section. On the other hand, of the 93 women with basal and/or stimulated FSH ≥ 10 IU/l, only seven (7.5%) started therapy sooner than would have been justified according to the criteria mentioned. Furthermore, the results of most ORTs may vary per cycle in the same woman, especially basal FSH and the CCCT (Scott *et al*., 1990; Kwee *et al*., 2004; Hendriks *et al*., 2005). In our study, the ORTs were only performed once per participant. It is not known whether repeating these tests would enhance their predictive value for spontaneous pregnancy.

We cannot exclude the possibility that our data underestimate the importance of the ORTs caused by an insufficient sample size. However, the comparison with the effect of the post-coital test makes it rather unlikely. The sample size was large enough to detect a statistically significant effect of the post-coital test of a size similar to other studies. In other words, if the effect of (one of) the ORTs would be close to that of the post-coital test, we would have detected it.

Comparison with other studies

Scott *et al*. (1993, 1995) evaluated basal FSH and the CCCT both prospectively and retrospectively in general subfertile populations of ovulatory and non-ovulatory women. They found that elevated basal and stimulated FSH values predicted low conception chances, spontaneous and treatment-related, if they exceeded a cut-off level determined in their own population. In contrast to Scott *et al*., we found some predictive value for basal, but none for stimulated FSH. Moreover, basal FSH only had predictive value if also low values were taken into account. One explanation for their different findings is the fact that they included women with irregular or absent menstrual cycles, who had possibly already entered menopausal transition or were climacteric. In our ovulatory population, we found no evidence that stimulated FSH or the complete CCCT (basal + stimulated FSH) have additional value over basal FSH alone. A small nested case–control study by Van Montfrans *et al*. (2000) showed no predictive value of basal FSH for spontaneous or treatment-related pregnancy using an arbitrary cut-off level for basal FSH. The prospective cohort study of Van Rooij *et al*. (2006) showed no linear relation between basal FSH, basal inhibin B, AFC or anti-Müllerian hormone and spontaneous pregnancy chances. They do not report on the existence of a non-linear relationship. Since most participants in their study started fertility treatment, the actual follow-up period in which spontaneous pregnancy chances could be assessed was short. A recent prospective study by Van der Steeg *et al*. (2007b) evaluated the use of basal FSH in 3519 subfertile ovulatory couples. They found a graphical relation between basal FSH values and time to spontaneous pregnancy, with an 'optimum' for basal FSH at 8 IU/l, comparable to our findings. Van der Steeg *et al*. used a simplified model describing no effect of lower basal FSH values, but a statistically significant linear increase in time to pregnancy with basal FSH values from 8 IU/l and up. The clinical relevance was limited: only 3.0% of their study population would change from the advice for expectant management to therapy or vice versa after addition of basal FSH to the Hunault model. In our population, only the complete range of basal FSH levels showed a statistically significant predictive value, demonstrating the influence of basal FSH levels <8 IU/l in our cohort. Possible explanations for these different findings are differences between size and characteristics of the study populations or the use of nine different FSH assays in the study of Van der Steeg *et al*.

Clinical implications

Our data do not support the routine use of ORTs in the basal subfertility evaluation of ovulatory couples for the prediction of spontaneous pregnancy chances. Clinical relevance of the addition of either basal FSH or AFC to the Hunault model is limited, especially when compared with the additional value of the post-coital test. In particular, we found no evidence that the addition of an ORT identifies a new group of couples with low pregnancy chances that would otherwise have been labelled incorrectly as candidates for expectant management. The lack of a clinically meaningful relation of either stimulated FSH or the full CCCT with spontaneous pregnancy chances in our ovulatory population is remarkable, showing that these tests have no value over basal FSH alone. Caution is needed when extrapolating our results to other populations. For instance, in populations of women of advanced age, the predictive value of ORTs could be different because of the expected higher prevalence of diminished ovarian reserve. Furthermore, our results cannot be readily applied to non-ovulatory women because of the different underlying pathology, such as polycystic ovarian syndrome or a climacteric cycle pattern.

Conclusions

Basal FSH and AFC have a statistically significant additional value to a validated prognostic model for spontaneous pregnancy. However, the clinical relevance of the addition of either of these tests is limited. Stimulated FSH, CCCT, basal and stimulated inhibin B offer no extra information on spontaneous pregnancy chances. We conclude that the ORTs tested in this study have no substantial added value to previously identified prognostic factors for spontaneous pregnancy, as reflected in the validated Hunault model. These results suggest that the ORTs evaluated should not be used routinely in the basal subfertility evaluation of ovulatory couples to predict spontaneous pregnancy chances.

Author's Role

M.L.H.: acquisition of data, analysis, interpretation, main author; H.G.: analysis and interpretation of data, critical review; V.F.: analysis and interpretation of data, critical review; A.B.: design, acquisition of data, critical review; E.M.A.R.: acquisition of

data, critical review; E.R.G.: acquisition of data, critical review; F.J.M.B.: interpretation of data, critical review; M.J.H.: design, acquisition of data, critical review; A.H.: design, acquisition of data, interpretation, critical review.

Acknowledgements

We thank Dr P.F.J. Donderwinkel for the inclusion of patients and the performance of ovarian reserve tests in the Martini Hospital. Parts of these data were presented at the 62nd Annual Meeting of the American Society for Reproductive Medicine, 2006, New Orleans, USA, and at the 23rd Annual Meeting of the European Society of Human Reproduction and Embryology, 2007, Lyons, France.

Funding

University Medical Center Groningen and Organon, the Netherlands.

References

Broekmans FJM, Faddy MJ, Scheffer GJ, Te Velde ER. Antral follicle counts are related to age at natural fertility loss and age at menopause. *Menopause* 2004;**11**:607–614.

Broekmans FJM, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update* 2006;**12**:685–718.

Collins JA, Burrows EA, Willan AR. The prognosis for live birth among untreated infertile couples. *Fertil Steril* 1995;**64**:22–28.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CWN, Habbema JDF. The prediction to conceive in subfertile couples. *Fertil Steril* 1994;**61**:44–52.

Haadsma ML, Bukman A, Groen H, Roeloffzen EMA, Groenewoud ER, Heineman MJ, Hoek A. The number of small antral follicles (2–6 mm) determines the outcome of endocrine ovarian reserve tests. *Hum Reprod* 2007;**22**:1925–1931.

Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**28**:361–387.

Hendriks DJ, Broekmans FJM, Bancsi LFJMM, De Jong FH, Looman CWN, Te Velde ER. Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve. *Hum Reprod* 2005;**20**:163–169.

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

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH, Te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019–2026.

Hunault CC, Laven JSE, Van Rooij IAJ, Eijkemans MJC, Te Velde ER, Habbema JDF. Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum Reprod* 2005;**20**:1636–1641.

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

Maheshwari A, Fowler P, Bhattacharya S. Assessment of ovarian reserve—should we perform tests of ovarian reserve routinely? *Hum Reprod* 2006;**21**:2729–2735.

Scheffer GJ, Broekmans FJM, Dorland M, Habbema JDF, Looman CWN, Te Velde ER. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. *Fertil Steril* 1999;**72**:845–851.

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

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.

Scott RT, Opsahl MS, Leonardi MR, Neall GS, Illions EH, Navot D. Life table analysis of pregnancy rates in a general infertility population relative to ovarian reserve and patient age. *Hum Reprod* 1995;**10**:1706–1710.

Steures P, Van der Steeg JW, Hompes PG, Habbema JDF, Eijkemans MJC, Broekmans FJM, Verhoeve HR, Bossuyt PMM, Van der Veen F, Mol BWJ. Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 2006;**368**:216–221.

Te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update* 2002;**8**:141–154.

Therneau TM, Grambsch PM. *Modelling Survival Data.* New York, USA: Springer, 2000.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGS, Broekmans FJM, Van Dessel HJHM, Bossuyt PMM, Van der Veen F, Mol BWJ. Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 2007a;**22**:536–542.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Broekmans FJM, Bouckaert PXJM, Bossuyt PMM, Van der Veen F, Mol BWJ. Predictive value and clinical impact of basal FSH in subfertile, ovulatory women. *J Clin Endocrinol Metab* 2007b;**92**:2163–2168.

Van Montfrans JM, Hoek A, Van Hooff MHA, De Koning CH, Tonch N, Lambalk CB. Predictive value of basal follicle-stimulating hormone concentrations in a general subfertile population. *Fertil Steril* 2000;**74**: 97–103.

Van Rooij IAJ, Broekmans FJM, Hunault CC, Scheffer GJ, Eijkemans MJC, De Jong FH, Themmen APN, Te Velde ER. Use of ovarian reserve tests for the prediction of ongoing pregnancy in couples with unexplained or mild male infertility. *Reprod Biomed Online* 2006;**12**:182–190.

Submitted on December 21, 2007; resubmitted on May 16, 2008; accepted on May 23, 2008