Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF

M.F.G. Verberg1,3, M.J.C. Eijkemans1,2, N.S. Macklon1, E.M.E.W. Heijnen1, B.C.J.M. Fauser1 and F.J. Broekmans1

1Department of Reproductive Medicine and Gynaecology, University Medical Centre Utrecht, Heidelberglaan 100, 3584 CS Utrecht, The Netherlands; 2Department of Public Health, Erasmus MC, University Medical Center, Dr. Molewaterplein 50, PO Box 1738, 3000 DR Rotterdam, The Netherlands
3Correspondence address. Tel: +31(0)302508373; Fax: +31(0)302505433; E-mail: m.f.g.verberg@umcutrecht.nl

BACKGROUND: Milder stimulation protocols are being developed to minimize adverse effects of ovarian stimulation in in vitro fertilization (IVF) programs. A drawback is the possibility of an increased rate of insufficient ovarian response. This study aimed to develop a prognostic model for the prediction of cycle cancellation due to insufficient response to mild stimulation. METHODS: A total of 174 IVF patients aged <38 years and with a body mass index (BMI) <28 Kg/m2 were treated with mild ovarian stimulation using a fixed daily dose (150 IU) of recombinant follicle-stimulating hormone (rFSH) from cycle day 5 and GnRH antagonist from the late follicular phase. In women with mono- or bifollicular growth (17%), the cycle was cancelled and the treatment was adjusted in a second treatment cycle by starting rFSH on cycle day 2. RESULTS: In a multivariable logistic regression analysis, duration of infertility, menstrual cycle length, secondary infertility and BMI were included in the prediction model. The area under the receiver-operating characteristics curve of the model was 0.69. A probability cut-off for cancellation of 0.3 yielded an expected sensitivity of 33% and specificity of 92%. Analysis of ovarian response in the subsequent treatment cycle showed an improved ovarian response and a significant reduction in the cancellation rate. CONCLUSIONS: With the presented model, it is possible to identify patients at risk for cycle cancellation, during mild ovarian stimulation, due to insufficient response. The contributing factors of the model suggest that ovarian aging and BMI are related to insufficient response to mild stimulation.

Keywords: mild ovarian stimulation; ovarian response; prediction models; IVF cycle cancellation

Introduction

As assisted reproductive science progresses, a shift in the focus of in vitro fertilization (IVF) is occurring from striving for maximizing instant success ‘at all costs’ to developing safer and more patient friendly protocols in which the risks of treatment are minimized while optimizing the chance of a singleton live birth (Edwards et al., 1996; Gerris, 2005). Ovarian stimulation is applied in IVF to generate multiple follicle growth in order to obtain an increased quantity of oocytes to compensate for inefficiency of the IVF procedure while maintaining the potential to select the best embryo (Fleming et al., 1990). Currently, this goal is usually achieved by a long gonadotrophin-releasing hormone (GnRH) agonist suppression protocol, in association with ovarian stimulation with high doses of exogenous follicle-stimulating hormone (FSH). Disadvantages of this approach are the high cost, complex stimulation protocols which take several weeks, physical and emotional discomfort, chances for complications and the essentially uncontrollable degree of ovarian response. The current trend of limiting the number of embryos to be transferred reduces the need for large numbers of oocytes. Moreover, there is increasing evidence of the detrimental effects of ovarian stimulation on corpus luteum function, endometrial receptivity and embryo quality (Valbuena et al., 2001; Bourgain and Devroey, 2003). As a consequence, mild ovarian stimulation protocols are being developed to minimize the adverse treatment effects of ovarian stimulation (Fauser et al., 1999). The introduction of GnRH antagonists into clinical practice and a greater understanding of the process of follicle recruitment and dominant follicle selection have led to new opportunities for developing mild stimulation protocols.

It has been shown that by interfering with the physiological decrease in FSH levels during the follicular phase, it is possible to override the selection of a single dominant follicle (van Santbrink et al., 1995; Hohmann et al., 2001). Both the degree and the duration of the FSH elevation will lead to an extension of the so called ‘FSH window’ that enables the development of several rather than just a single dominant follicle.
(Zelevnik et al., 1985; Schipper et al., 1998; de Jong et al., 2000). Indeed a slight but extended elevation of FSH levels during the mid to late follicular phase has been shown to be sufficient for growth of a modest number of dominant follicles (Schipper et al., 1998; Hohmann et al., 2001). Yet, there appears to be an individual variability in the optimal moment for initiating exogenous FSH supplementation. Hohmann et al. (2001) observed no difference in multifollicular growth when daily FSH administration was started on day 3 or 5, although a tendency toward lower numbers of dominant follicles was seen when cycle day 7 was chosen to initiate FSH injections.

With the use of GnRH antagonists to prevent a premature luteinizing hormone rise, the IVF treatment cycle can commence at a point in the course of a spontaneous menstrual cycle where the recruitment of a cohort of antral follicles has already been established (Fauser and van Heusden, 1997; Macklon and Fauser, 2000). This approach enables limiting the use of exogenous FSH in order to extend the FSH window, allowing multiple dominant follicle development to take place. As a consequence, the number of treatment days and the total amount of exogenous FSH required is substantially reduced (de Jong et al., 2000; Hohmann et al., 2003; Heijnjen et al., 2007). A potential drawback of mild stimulation is a decrease in ovarian response compared with conventional stimulation, leading to higher cancellation rates (Fauser et al., 1999; Hohmann et al., 2003). Although low numbers of oocytes appear to be related to good outcomes in mild stimulation (Hohmann et al., 2003), cancellations should be prevented to optimize the benefit of mild stimulation. The purpose of this study was the prediction of which mild stimulation cycles are likely to be cancelled due to insufficient ovarian response. The development of methods to identify women who may benefit from an earlier start of exogenous FSH may reduce the number of cancelled cycles and improve the efficacy of the mild stimulation protocol.

Materials and Methods

Study design

Data were derived from the mild stimulation arm of a randomized controlled trial on effectiveness of IVF treatment strategies (Heijnjen et al., 2007). The study was approved by the local ethics review board of both participating centers. In this study, infertile patients, with a regular indication for IVF or intracytoplasmic sperm injection (ICSI) and who attended the Erasmus Medical Centre (Rotterdam, the Netherlands) or the University Medical Centre Utrecht (Utrecht, The Netherlands), were invited to participate. Participants were <38 years of age and had a regular menstrual cycle (25–35 days) and a body mass index (BMI) between 18–28 kg/m². Couples who had been previously treated with IVF were excluded. Study design and clinical outcomes of the RCT have been reported recently (Heijnjen et al., 2007).

Patients in the mild stimulation arm were treated with a fixed daily starting dose of 150 IU recombinant FSH (rFSH) (Gonal-F®: Serono Benelux B.V., Amsterdam, The Netherlands; or Puregon®: N.V. Organon, Oss, The Netherlands) s.c., initiated on the fifth cycle day (CD 5 protocol). The dose of exogenous FSH was not adjusted during the stimulation. GnRH antagonist (ganirelix, Orgalutran®: N.V. Organon, 0.25 mg/day; or cetorelix, Cetrotide®: Serono Benelux, 0.25 mg/day) was administered s.c. from the day that at least one follicle attained a diameter ≥14 mm (Hohmann et al., 2003). Human chorionic gonadotrophin (hCG) (Profasi®: Serono Benelux B.V.; or Pregnyl®: N.V. Organon) 10 000 IU s.c. was administered as a single bolus injection to induce final oocyte maturation, when the largest follicle had reached at least 18 mm in diameter and at least one additional follicle ≥15 mm was observed. Oocyte retrieval and fertilization ‘in vitro’ was performed according to standard procedures as described previously (Kastrop et al., 1999; Huisman et al., 2000). Single embryo transfer of the resulting best quality embryo was performed on day 3 or 4 after oocyte retrieval. Standard luteal phase support in the form of intravaginal progesterone (Progatest®: N.V. Organon) 600 mg/day was given from the day of oocyte retrieval until a urine pregnancy test was performed 18 days later.

Insufficient ovarian response resulting in cancellation of the cycle was defined as the development of less than three follicles >12 mm. In these patients, exogenous FSH was initiated on cycle day 2 (CD 2 protocol) in a subsequent treatment cycle while the daily dosage remained unchanged. Cycles at risk for ovarian hyperstimulation syndrome (OHSS), defined as more than 20 follicles with a diameter >10 mm or estradiol concentrations >15,400 pmol/l were also cancelled before hCG injection.

Data analysis

In order to identify a prior predictors of cancellation cycles due to insufficient response in a mild stimulation protocol, female age, previous pregnancy, previous childbirth, cause of infertility, menstrual cycle length, IVF or ICSI treatment and BMI were compared between patients where ovarian stimulation was cancelled and patients who had a sufficient ovarian response to proceed to oocyte pick-up in their first treatment cycle. For this analysis, cycles cancelled due to increased risk of OHSS were considered as good responders and included in the analysis. Cycles cancelled due to premature luteinization were excluded from the analysis. Multivariable logistic regression analysis was performed with a backward elimination procedure, a P-value <0.1 was used as a criterion for exclusion. The predictive ability of the model was assessed by determining the area under the receiver operating characteristics (ROC) curve (AUC).

To assess the amount of overfitting of the created model, internal validation was performed with bootstrapping, a statistical technique to create comparable populations. We bootstrapped 200 times. In each of these 200 new data sets, the same multivariable logistic regression analysis with backward elimination was performed, and the resulting model was tested on the original data. In this way the amount of overfitting can be assessed and expressed as a shrinkage factor. The shrinkage factor should be taken into account when applying the model in clinical practice (Van Houwelingen and le Cessie, 1990; Harrell et al., 1996).

To study whether patients who were cancelled due to insufficient ovarian response in their first treatment cycle with the CD 5 protocol presented with improved ovarian response when stimulation in the subsequent cycle was started at CD 2, a ‘within patient comparison’ was performed. In this analysis, the ovarian response of a second treatment is compared with the response in the first treatment cycle of the same individual. To compare both cycles, duration of stimulation, total rFSH needed and number of developed follicles and oocytes of patients who were cancelled for insufficient ovarian response in their first treatment cycle were included in the analysis.

Comparisons of outcome measures between the groups were performed using the t-test for continuous data and the χ²-test for binary variables unless stated otherwise. Within patient comparison was done by a paired t-test. Analyses were performed using SPSS
Results
Of the 174 first cycles started, 39 (22%) ended in a cancellation: 30 (17%) due to an insufficient response and 9 (5%) for other reasons (Fig. 1). Univariable analysis of patient characteristics is shown in Table I. A significantly shorter menstrual cycle length (the number of days of an average menstrual cycle in the previous year as indicated by the patient) (28.2 versus 27.5 days; \(P = 0.045\)) and longer duration of infertility (4.4 versus 3.6 years; \(P = 0.022\)) were observed in patients with an insufficient response. The number of treatment days and medication used were also compared between cancellation for insufficient ovarian response and patients who did continue for follicle aspiration. As expected, there was a significant difference in treatment days and medication used in these cycles (\(P\)-values all \(<0.001\)) (Table II).

In the multivariable analysis, the variables, duration of infertility, menstrual cycle length, primary or secondary infertility and BMI, were selected into the prediction model for cancellation during the mild stimulation protocol. A longer duration of infertility, short menstrual cycle length, secondary infertility and higher BMI were found to be associated with an insufficient ovarian response. The predictive ability of the model measured by the area under the ROC curve was 0.69 (95% confidence interval [CI] 0.58–0.79) (Table III). Internal validation by bootstrapping showed a shrinkage factor of 0.58. This means that the final model is overfitted to our data and the predictive probabilities of the model on external data will be less. The model will give the most reliable predictions when all regression coefficients are on average 42% smaller in absolute size (Table III). A probability cut-off for cancellation due to insufficient response of 0.3 yielded an expected sensitivity of 33%, specificity of 92%, and positive predictive value of 48% on our own data. A probability cut-off of 0.15 yielded an expected sensitivity of 77%, specificity of 54% and positive predictive value of 26%. Table IV shows the validity of the model with a range of cut-off values chosen in the area of the AUC curve of the model with the most discriminative power.

As expected, the within patient comparison between the first and second treatment cycles of patients who were cancelled for insufficient response in their first treatment cycle showed that a change in treatment protocol from CD 5 to CD 2 resulted in a clear improvement in ovarian response. Of the 27 patients whose first treatment cycle was cancelled and who proceeded with a changed protocol in the second cycle, 3 (11%) had to be cancelled again.

Discussion
The aim of the present study was to explore whether it is possible to identify a subgroup of patients at risk for cancellation due to insufficient ovarian response in mild ovarian stimulation for IVF starting exogenous FSH on cycle day 5. Our study confirmed the previous finding of a relatively high cancellation rate (17%) in the mild stimulation protocol (Hohmann et al., 2003). To increase the benefit of mild stimulation we analysed characteristics of patients with an insufficient ovarian response. According to pre-treatment variables we developed a prognostic model to identify patients at risk for cancellation prior to the start of ovarian stimulation. The predictors in the model were a longer duration of infertility, a shorter menstrual cycle length, secondary infertility and a higher BMI.

The finding of an association between a shorter cycle length and insufficient ovarian response may be the consequence of a shortened follicular phase in these patients where dominant follicle selection may have occurred prior to the start of the exogenous FSH (Klein et al., 2002; van Zonneveld et al., 2003). As a consequence, involvement of the non-dominant follicles in ongoing growth by exogenous FSH becomes impossible. Also, cycle shortening may be a subtle first sign.
of advanced ovarian aging and as such be related to a small cohort size. As only patients with a regular cycle length and below 38 years of age were included for this study, this phenomenon may become even more evident in the general population.

A negative association between BMI and ovarian response has been observed previously (Wittemer et al., 2000; Nichols et al., 2003; Fedorcsak et al., 2004). Obese patients usually require significantly higher doses of gonadotrophin and a longer duration of stimulation (Dechaud et al., 1998). However, in a prospective study comparing predictive factors of ovarian response in IVF, BMI did not associate with the number of follicles or the number of retrieved oocytes (Popovic-Todorovic et al., 2003).

Longer duration of infertility has previously been recognized as an important negative prognostic factor affecting the chance of natural conception, particularly in unexplained infertility (Hull et al., 1985). This might be the consequence of subtle, undiagnosed disorders related to a diminished ovarian reserve. These disorders might normally be overcome by fierce stimulation in more conventional stimulation protocols, but are noticed in natural conception and mild stimulation.

The observed association between ovarian response and secondary infertility has not been described previously. In our opinion, there is no straight-forward biological explanation for this relationship. Its appearance in the model could be the result of the relatively small sample size used to build the model and might explain the relatively large shrinkage factor that was found in the internal validity analysis.

The only previous study regarding the current mild stimulation protocol with a similar analysis showed a significant difference in age and baseline FSH between cancelled patients and those who met the criteria for oocyte retrieval (Hohmann et al., 2003). The current study could not confirm the finding of female age as an important predictor in spite of the larger sample size. This might be partly the result of the age restriction in the inclusion criteria for the study. Additionally, both the shorter cycle length and the longer duration of infertility appearing in the current model might be related to ovarian aging resulting in age being removed in multivariable analysis.

The only previous study regarding the current mild stimulation protocol with a similar analysis showed a significant difference in age and baseline FSH between cancelled patients and those who met the criteria for oocyte retrieval (Hohmann et al., 2003). The current study could not confirm the finding of female age as an important predictor in spite of the larger sample size. This might be partly the result of the age restriction in the inclusion criteria for the study. Additionally, both the shorter cycle length and the longer duration of infertility appearing in the current model might be related to ovarian aging resulting in age being removed in multivariable analysis.

### Table I: Univariable analysis of patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Cancellation due to insufficient ovarian response (n = 30)</th>
<th>Sufficient ovarian response (n = 140)</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age (years)</td>
<td>33.1 ± 3.0</td>
<td>32.7 ± 3.1</td>
<td>0.45 (−1.68, 0.77)</td>
</tr>
<tr>
<td>Primary infertility (%)</td>
<td>70%</td>
<td>82%</td>
<td>−</td>
</tr>
<tr>
<td>Cause of infertility (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal</td>
<td>10%</td>
<td>16%</td>
<td>−</td>
</tr>
<tr>
<td>Male</td>
<td>43%</td>
<td>54%</td>
<td>−</td>
</tr>
<tr>
<td>Unexplained</td>
<td>37%</td>
<td>26%</td>
<td>−</td>
</tr>
<tr>
<td>Other</td>
<td>10%</td>
<td>4%</td>
<td>−</td>
</tr>
<tr>
<td>Duration of infertility (years)</td>
<td>4.4 ± 2.4</td>
<td>3.6 ± 1.7</td>
<td>0.85 (−1.57, −0.13)*</td>
</tr>
<tr>
<td>Cycle length (days)</td>
<td>27.5 ± 1.5</td>
<td>28.2 ± 1.9</td>
<td>0.74 (0.15, 1.47)*</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.5 ± 3.0</td>
<td>23.0 ± 2.5</td>
<td>0.46 (−1.49, 0.58)</td>
</tr>
</tbody>
</table>

Values are means ± SD.
* Differences are statistically significant (P-value < 0.05).
* Including five cycles that were cancelled as a consequence of increased risk for OHSS.

### Table II: Univariable analysis of cycle characteristics

<table>
<thead>
<tr>
<th></th>
<th>Cancellation due to insufficient ovarian response (n = 30)</th>
<th>Sufficient ovarian response (n = 135)</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of stimulation (days)</td>
<td>5.5 ± 1.7</td>
<td>8.0 ± 1.6</td>
<td>2.47 (1.84, 3.01)*</td>
</tr>
<tr>
<td>Total rFSH used (IU)</td>
<td>830 ± 248</td>
<td>1200 ± 256</td>
<td>370 (268, 472)*</td>
</tr>
<tr>
<td>GnRH antagonist administered (days)</td>
<td>1.4 ± 1.2</td>
<td>3.7 ± 1.4</td>
<td>2.29 (1.72, 2.84)*</td>
</tr>
</tbody>
</table>

Values are means (± SD).
* Differences are statistically significant (P-value < 0.05).
* Insufficient ovarian response was defined as the development of less than three dominant follicles (diameter > 12 mm).

### Table III: Multivariable analysis for cancellations due to insufficient response in the mild CD 5 stimulation protocol; the ability of the model measured by the area under the ROC curve was 0.69 (95% CI: 0.58–0.79)

<table>
<thead>
<tr>
<th></th>
<th>P-value</th>
<th>Odds Ratio (95% CI)</th>
<th>Cumulative AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of infertility</td>
<td>0.033</td>
<td>1.24 (1.02, 1.50)</td>
<td>0.60</td>
</tr>
<tr>
<td>Menstrual cycle length</td>
<td>0.034</td>
<td>0.75 (0.59, 0.98)</td>
<td>0.67</td>
</tr>
<tr>
<td>Secondary infertility</td>
<td>0.13</td>
<td>2.08 (0.82, 5.27)</td>
<td>0.68</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>0.26</td>
<td>1.10 (0.93, 1.29)</td>
<td>0.69</td>
</tr>
</tbody>
</table>

*Predicted probability of cancellation due to insufficient response: 1/(1 + Exp(−(0.608 + 0.126 duration − 0.158 [cycle length] + 0.423 [secondary infertility (yes = 1, no = 0)] + 0.055*BMI))).
Analysis of ovarian response in a subsequent treatment cycle with the CD 2 protocol for patients who previously presented with an insufficient ovarian response showed that these were likely to meet the criteria for oocyte retrieval. These results support the hypothesis that in a mild stimulation protocol, insufficient ovarian response is a consequence of suboptimal ovarian stimulation for a specific group of patients and can be overcome by the early commencement of exogenous stimulation. However, it is likely that at least part of the improved ovarian response is the result of the principle of regression toward the mean. This is a principle stating that of related measurements, and selecting those where the first measurement is either higher or lower than the average, the expected value of the second is closer to the mean than the observed value of the first (Davis, 1976). Prospective randomized studies are needed to establish which part of the improved ovarian response should be ascribed to the change in the hormonal stimulation schedule.

These data indicate that the CD 2 protocol is likely to improve chances for patients who would be cancelled in the CD 5 protocol due to insufficient ovarian response. Because the prediction model is based on a priori parameters, patients at risk for cancellation can be identified prior to the start of the treatment. The overall cancellation rate for insufficient response is therefore likely to be reduced if these patients are treated with the CD 2 protocol instead of the CD 5 protocol.

Although the area under the ROC curve was modest, the model has, in our opinion, the ability to predict sufficient numbers of cancellations to reduce the remaining number of cancellations to an acceptable level. At a cut-off level of 0.30, the model will predict 10 cancellations correctly (sensitivity 33% (95% CI 16.5–50.2)). According to our CD 2-started second cycles, 1 out of these 10 would still be cancelled (11%). This means that with a model based treatment adaptation the percentage cancels would be lowered from 17 to 12% (21/174). This reduction would equal the proportional difference in cancellation rates between the days 2 and 5 arm of the study by Hohmann et al. (2003), and as such may be considered clinically useful. Due to the high specificity at the cut-off level (92% (95% CI 87.7–96.6)), the number of patients that will be unnecessarily treated with the CD 2 protocol is limited, so that the majority of cases can therefore still benefit from the advantages of the mild stimulation day 5 protocol. Still, one should be aware that the model is developed on a modest number of patients leading to the relative broad confidence interval for sensitivity (16.5–50.2%).

Overall, the clinical use of the prediction model may render mild ovarian stimulation more patient tailored. The concept of ‘cycle cancellation’ should be viewed in the context of just a few days of medication in the current mild stimulation approach compared with the conventional approach. In such stimulation protocols (GnRH agonist long protocol, sometimes proceeded by oral contraceptives and followed by extended ovarian stimulation), the time elapsed before an insufficient response can be identified is much longer, resulting in an increased waste of medication and the delay of at least two menstrual cycles. The cancellation rate during mild stimulation should also be balanced against the extended gain (for multiple cycles, if needed) of a later start of stimulation in the great majority of women resulting in fewer injections and reduced patient discomfort and cost.

In conclusion, midfollicular initiation of rFSH in combination with a GnRH antagonist leads to a mild ovarian stimulation protocol, but yields a distinct risk of cancellation due to insufficient response. We developed a model that could predict 33% of the cancellations (sensitivity) with a false positive rate of 8% on our own data and so equalize the cancellation rate for insufficient response to that observed in a standard GnRH antagonist ovarian stimulation protocol for IVF. Insufficient response in mild stimulation may be related to ovarian aging and increased BMI. After external validation, the model might be used to identify patients at risk for insufficient ovarian response prior to the start of the treatment cycle. Treatment with ovarian stimulation initiated in the early follicular phase in these patients may reduce the number of cancelled cycles and therefore improve the efficacy of the mild stimulation protocol. Further prospective randomized trials should evaluate the clinical use of adjusting the starting day of ovarian stimulation based on the current prediction model.

### Acknowledgements

This study was funded by ZonMw (The Netherlands), program Doelmatigheidsonderzoek (grant number 945-12-010).

### References


Hohmann FP, Macklon NS, Fauser BC. A randomized comparison of two ovarian stimulation protocols with gonadotropin-releasing hormone (GnRH) antagonist cotreatment for in vitro fertilization commencing the "follicle-stimulating hormone window" combined with the gonadotropin-releasing hormone (GnRH) antagonist protocol. *J Clin Endocrinol Metab* 2003;88:166–173.


Scherper I, Hop WC, Fauser BC. The follicle-stimulating hormone (FSH) threshold/window concept examined by different interventions with exogenous FSH during the follicular phase of the normal menstrual cycle: duration, rather than magnitude, of FSH increase affects follicle development. *J Clin Endocrinol Metab* 1998;83:1292–1298.

Valbuena D, Martin J, de Pablo JL et al. Increasing levels of estradiol are deleterious to embryonic implantation because they directly affect the embryo. *Fertil Steril* 2001;76:962–968.


