Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples*

J.W. van der Steeg1,2,3,8, P.Steures1,2,3, M.J.C. Eijkemans1, J.D.F. Habbema1, P.G.A. Hompes3, F.J. Broekmans4, H.J.H.M. van Dessel5, P.M.M. Bossuyt6, F. van der Veen2 and B.W.J. Mol2,7 on behalf of CECERM study group†

1Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, 2Center for Reproductive Medicine, Academic Medical Center, Amsterdam, 3Department of Obstetrics & Gynaecology, Vrije Universiteit Medical Center, Amsterdam, 4Division of Reproductive Medicine, Department of Perinatology and Gynaecology, University Medical Center, Utrecht, 5Department of Obstetrics & Gynaecology, TweeSteden Ziekenhuis, Tilburg, 6Department of Clinical Epidemiology and Biostatistics, Academic Medical Center and 7Department of Obstetrics & Gynaecology, Máxima Medical Center, Veldhoven, The Netherlands

*To whom correspondence should be addressed at: Academic Medical Center, Center of Reproductive Medicine, Room H4-213, Department of Obstetrics/Gynaecology, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. E-mail: j.w.vandersteeg@amc.uva.nl


BACKGROUND: Prediction models for spontaneous pregnancy may be useful tools to select subfertile couples that have good fertility prospects and should therefore be counselled for expectant management. We assessed the accuracy of a recently published prediction model for spontaneous pregnancy in a large prospective validation study.

METHODS: In 38 centres, we studied a consecutive cohort of subfertile couples, referred for an infertility work-up. Patients had a regular menstrual cycle, patent tubes and a total motile sperm count (TMC) >3 × 10^6. After the infertility work-up had been completed, we used a prediction model to calculate the chance of a spontaneous ongoing pregnancy (www.freya.nl/probability.php). The primary end-point was time until the occurrence of a spontaneous ongoing pregnancy within 1 year. The performance of the pregnancy prediction model was assessed with calibration, which is the comparison of predicted and observed ongoing pregnancy rates for groups of patients and discrimination.

RESULTS: We included 3021 couples of whom 543 (18%) had a spontaneous ongoing pregnancy, 57 (2%) a non-successful pregnancy, 1316 (44%) started treatment, 825 (27%) neither started treatment nor became pregnant and 280 (9%) were lost to follow-up. Calibration of the prediction model was almost perfect. In the 977 couples (32%) with a calculated probability between 30 and 40%, the observed cumulative pregnancy rate at 12 months was 30%, and in 611 couples (20%) with a probability of ≥40%, this was 46%. The discriminative capacity was similar to the one in
which the model was developed (c-statistic 0.59). CONCLUSIONS: As the chance of a spontaneous ongoing pregnancy among subfertile couples can be accurately calculated, this prediction model can be used as an essential tool for clinical decision-making and in counselling patients. The use of the prediction model may help to prevent unnecessary treatment.

Key words: model/spontaneous pregnancy/prognosis/subfertility

Introduction
Infertility or involuntary childlessness is one of life’s great catastrophes, which happens to approximately one of ten couples (Gibbons, 1911; Hull et al., 1985; Gnoth et al., 2003; Wang et al., 2003). Most of these couples seek medical help (Evers, 2002). Assessment of the fertility potential in an infertility work-up is then the first step (National Collaborating Centre for Women’s and Children’s Health, 2004). After the completion of the infertility work-up, it is essential to distinguish subfertile couples in whom prognosis of spontaneous pregnancy is poor and fertility treatment is mandatory from subfertile couples who still have a good prognosis to conceive spontaneously (te Velde and Cohlen, 1999). Prediction models for spontaneous pregnancy may be useful tools to achieve this (Wasson et al., 1985).

In the last decade, three prediction models for the prediction of spontaneous pregnancy in subfertile couples have been published (Eimers et al., 1994b; Collins, et al. 1995; Snick, et al. 1997). In 1994, Eimers et al. developed a model that was based on >900 subfertile couples who visited a university fertility centre in the Netherlands before the era of IVF. In 1995, Collins et al. developed a model that was based on tertiary care subfertile couples who visited one of the 11 participating IVF centres in Canada. Finally, in 1997, Snick et al. developed a model that was based on data derived from a secondary care fertility centre, on the Walcheren peninsula in the Netherlands.

To broaden the empirical basis of these pregnancy predictions models, the data of the three prediction models were pooled and integrated in a synthesis model (Hunault et al., 2004). This synthesis model is a pregnancy prediction model that contains the variables female age, duration of subfertility, subfertility being primary or secondary, semen motility and referral status and is available in a version with and without the ability of the pregnancy prediction model to differentiate spontaneously (te Velde and Cohlen, 1999). The question is, therefore, why these prediction models for spontaneous pregnancy may be useful tools to achieve this (Wasson et al., 1985).

At the moment, none of the above-described pregnancy prediction models have been implemented in fertility guidelines or overviews [Bagshawe and Taylor, 2003; Boston (MA), 2003; National Collaborating Centre for Women’s and Children’s Health, 2004; Bloomington (MN): Institute for Clinical Systems Improvement (ICSI), 2004], even though it has been argued that couples must be informed about their chances of conceiving spontaneously before considering any treatment (te Velde and Cohlen, 1999). The question is, therefore, why these prediction models have not yet been recommended in guidelines. One important reason may be the fact that the general applicability of these pregnancy prediction models has not been assessed in prospective validation studies in a wide variety of settings (Wasson et al., 1985). This can be done by external validation, an essential and final step in the development of prediction models (Wasson et al., 1985; Bleeker, et al. 2003), because prediction models tend to be overoptimistic when applied in other populations than the one in which they were developed (Stolwijk et al., 1996, 1998).

The aim of this study was to validate the synthesis prediction model for spontaneous pregnancy in prospectively collected data from an external population. We also assessed the ability of the pregnancy prediction model to differentiate couples with a good chance to conceive from couples in whom treatment should be started because their prospects are too poor.

Materials and methods
The study was designed as a prospective cohort study performed in 38 hospitals in the Netherlands. The local ethics committee of each participating centre gave Institutional Review Board approval for this study. Between January 2002 and February 2004, we included consecutive subfertile couples who had not been evaluated previously for subfertility. All couples were referred by their general practitioner and underwent an infertility work-up, consisting of a fertility history, semen analysis, PCT, assessment of ovulation and assessment of the Fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (2004). Couples referred by a gynaecologist were not included in this study, because these couples were referred for fertility treatment rather than for evaluating their fertility potential.

The duration of subfertility was defined as the period between the time the couple had an active child wish and the moment at which the infertility work-up was completed. If the couple had a previous pregnancy that had not resulted in a live birth, the duration of subfertility was defined as the period between the renewed active child wish after this pregnancy and the moment at which the infertility work-up was completed. Female age was calculated at the time the infertility work-up ended. At the time of the study, no treatment was offered to women >40 years. Subfertility was considered to be secondary if a woman had conceived in the current or in a prior partnership, regardless of the pregnancy outcome.

The menstrual cycle was considered regular if the duration of the cycle was between 23 and 35 days, with an intercycle variation of <8 days. Presence of ovulation was confirmed by means of a basal body temperature (BBT) chart, a mid-luteal serum progesterone, or by sonographic monitoring of the cycle. Women with an ovulation disorder, i.e. an irregular menstrual cycle or anovulation, were not included in the study.

Semen analysis was performed at least once. Couples in whom the man had a total motile sperm count (TMC) <3 × 10^6 were excluded from the analysis.

At least one PCT was performed during the infertility work-up (Eimers et al., 1994a; Snick et al., 1997; Glazener et al., 2000; van der Steeg et al., 2004). The PCT could be planned based on the BBT and...
cycle length or on repetitive ultrasound findings. The PCT was judged to be normal if at least one progressively motile spermatozoon was seen in one of five high-power fields at ×400 magnification. All other PCT results were considered to be abnormal.

Tubal pathology was assessed by a Chlamydia Antibody Test (CAT) or directly assessed by hysterosalpingography (HSG) or laparoscopy. In case of a positive CAT, the tubal status was subsequently evaluated with HSG or laparoscopy, whereas in cases of a negative CAT, tubal pathology was considered absent (Mol et al., 1997). The CAT could be tested with immune fluorescence technique (IF) or with enzyme immune assays (EIA) (BioMerieux, Paris, France; Medac GmbH, Wedel, Germany; Savyon Diagnostics, France). The CAT was considered to be positive for IF if the titre was >1:16 and for ELISA if the level was >1.1. Women with one- or two-sided tubal pathology were excluded from further follow-up and analysis.

After the completion of the infertility work-up, we calculated a probability of spontaneous pregnancy within 1 year. To do so, we used the pregnancy prediction model developed by Hunault et al. (2004). This model is based on a Cox regression model developed to predict spontaneous pregnancy within 1 year, leading to live birth, fitted to the data used in the construction of previous prediction models by Eimers, Collins and Snick (Eimers et al., 1994b; Collins et al., 1995; Snick et al., 1997) and as such is called the ‘synthesis’ model. It includes six prognostic variables: female age, duration of subfertility, female subfertility being primary or secondary, percentage motile spermatozoa of the first semen analysis, result of the PCT and referral status (being referred by general practitioner or gynaecologist). Each variable is converted in a point score. The total point score of each couple corresponds to a prognosis of a spontaneous ongoing pregnancy. The computer model can be used with following URL: http://www.frey.nl/probability.php. The synthesis prediction model is available in two variants; one includes the PCT, whereas the other does not include the PCT. In this manuscript, we evaluate the model without the PCT, which we refer to as the prediction model for spontaneous pregnancy. The results of the model with the PCT are presented in the discussion.

Couples in whom the prognosis for spontaneous pregnancy was ≥40% within 12 months were counselled for expectant management for a period of at least 6 months. After 6 months of expectant management, it was up to the couples to decide whether to start treatment or to wait for a longer period. Couples with a prognosis <40% were counselled for treatment according to the national fertility guidelines (Dutch Society of Obstetrics and Gynaecology, 1998, 2000).

Follow-up started at the completion of the infertility work-up and ended after 12 months. Primary end-point in this study was time to spontaneous conception resulting in an ongoing pregnancy. The first day of that menstrual cycle was considered to mark the end of time until spontaneous conception. Spontaneous ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal sonography at a gestational age of at least 12 weeks, resulting from a treatment-independent conception. Time to pregnancy was censored at the moment treatment had been started within 12 months after counselling or at the last date of contact during follow-up, when the couple had no ongoing pregnancy. In case of a miscarriage or ectopic pregnancy, follow-up continued. The time to pregnancy in these couples was not censored at the time of the unsuccessful pregnancy. For all couples who were lost to follow-up, the general practitioner was sent a questionnaire and asked about the fertility status of the couple.

Analysis

Missing data of the predictive variables were imputed (‘filled in’), because deleting them would lead to a loss of statistical power in multivariable analysis and—more seriously—potentially biased results (Little and Rubin, 1987; Schafer, 1997). We generated a single imputed dataset, using the first step of the ‘aRegImpute’ multiple imputation function in Splus 6.0. This is an efficient implementation of Bayesian multiple imputation, a recommended state of the art method (Schafer and Graham, 2002). The predictors were female age, duration of subfertility, previous pregnancy, semen motility, PCT and the referral status. We generated an imputed dataset, using the ‘aRegImpute’ imputation function in S-plus® 6.0.

We evaluated the calibration of the prediction model for spontaneous pregnancy (Hunault et al., 2004). Calibration is the degree of comparison between observed and predicted event rates for groups of patients (Altman and Royston, 2000). Calibration was assessed by comparing the mean predicted probability with the mean observed fraction of ongoing pregnancies at 12 months, in 10 subgroups. For this purpose, the cohort was split into 10 groups based on the deciles of the calculated probabilities. Per group, the mean predicted probability as well as the mean observed fraction, calculated with the Kaplan Meier method, was calculated. For all groups, we plotted the predicted and observed means in a calibration plot. We calculated the slope and its confidence interval (CI) for the regression curve of the plot. In case of a perfect calibration, all predictions and observations would be located on the line of equality (X = Y), and the slope would be unity (slope = 1) with a P-value >0.05 (not significantly different from 1).

Discrimination is a measurement of the ability to distinguish between patients who do and do not experience the event of interest, in this case getting pregnant without treatment (Altman and Royston, 2000). Discriminative capacity was assessed with receiver operation characteristic (ROC) analysis. We also calculated the c-statistic. This statistic is comparable to the area under the ROC-curves (AUC) but corrects for the fact that some couples were censored before follow-up of 12 months (Harrell et al., 1996).

Calculations were performed with SPSS® 12.0 (SPSS, Chicago, IL) and S-plus® 6.0 (MathSoft, Seattle, WA) programs.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

In the period between 1 January 2002 and 1 February 2004, we registered 5591 subfertile couples. Imputation was done on all patients who had at most two missing values in the six core prognosticators for spontaneous pregnancy. In total, 4.3% data points were missing and subsequently imputed. Of these 5591 patients, 948 had a severe male factor, 311 had two-sided tubal pathology and 1311 had one-sided tubal pathology, leaving 3021 couples for inclusion. Baseline characteristics of these couples are summarized in Table I.

In 1917 couples (64%), the estimated probability of a spontaneous pregnancy was <40%, and possible treatment was discussed with the couples. In this group, 17% became pregnant without treatment. In 1104 couples (36%), the probability of a spontaneous pregnancy was ≥40% and expectant management for at least 6 months was recommended. In this group, 52% of the couples fulfilled the 6-month period of expectant management, whereas 30% were treated before 6 months.
The follow-up status of all patients at 12 months is shown in Figure 1. We completed follow-up for 2741 couples (91%). Of the 3021 couples, 543 (18%) had a spontaneous ongoing pregnancy within 1 year, including 10 multiple pregnancies (1.8%). Unsuccessful pregnancies occurred in 57 couples (1.9%), of which 53 resulted in a miscarriage and four resulted in an ectopic pregnancy, i.e. 9.5 and 0.7% of all pregnancies, respectively. Within 12 months, 1316 (44%) of all couples started treatment, whereas 825 (27%) did neither start treatment nor became pregnant within the follow-up period.

The mean probability of a spontaneous pregnancy as calculated with the prediction model was 0.32 (5th and 95th percentiles: 0.16 and 0.55). The overall cumulative fraction of untreated couples with a spontaneous pregnancy at 12 months was 29.5% (95% CI: 27–32%) (Figure 2).

Results of the calibration analysis of the prediction model for spontaneous pregnancy are shown in Figure 3. This figure shows the association between the mean calculated probability of spontaneous ongoing pregnancy and the mean observed fraction of spontaneous ongoing pregnancies for each of the 10 decile groups. The prediction model for spontaneous pregnancy had a good calibration, with a slope of 0.82 (95% CI 0.6–1.0, P-value 0.08). In the 977 couples (32%) with a calculated probability between 30 and 40%, the observed cumulative pregnancy rate at 12 months was 30%, and in 611 couples (20%) with a probability of ≥40%, this was 46%.

For an assessment of the discriminative capacity, we calculated a c-statistic of 0.59 for the prediction model for spontaneous pregnancy, for both the version with and the version without the PCT. The discriminative capacity was similar to the one on which the model was developed, where the c-statistic had values of 0.63 and 0.59 for the version with and without the PCT, respectively.

**Discussion**

Our study was designed to validate the recently developed synthesis prediction model for spontaneous pregnancy in subfertile couples in routine clinical practice. For a period of 2 years, 38 hospitals used this prediction model to counsel subfertile couples for expectant or active management according to their

---

**Table I.** Baseline characteristics of the 3021 included subfertile couples

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean/Median</th>
<th>5th–95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age (years)</td>
<td>32.2</td>
<td>25–39</td>
</tr>
<tr>
<td>Male age (years)</td>
<td>34.8</td>
<td>27–44</td>
</tr>
<tr>
<td>Duration of subfertility (years)*</td>
<td>1.7</td>
<td>1.0–3.9</td>
</tr>
<tr>
<td>Subfertility, primary (n)(%)</td>
<td>2013</td>
<td>67</td>
</tr>
<tr>
<td>Semen analysis—TMC (10⁶)*</td>
<td>55.2</td>
<td>4.0–315</td>
</tr>
<tr>
<td>PCT, normal (n)(%)</td>
<td>2023</td>
<td>67</td>
</tr>
<tr>
<td>Cycle length (days)</td>
<td>28.1</td>
<td>24–33</td>
</tr>
<tr>
<td>FSH (U/l)*</td>
<td>6.4</td>
<td>3.5–12</td>
</tr>
<tr>
<td>BMI (kg/m²)*</td>
<td>23.0</td>
<td>19–34</td>
</tr>
</tbody>
</table>

BMI, body mass index; FSH, follicular stimulating hormone; PCT, postcoital test; TMC, total motile sperm count.

*Value is the median.

---

**Figure 1.** Study profile.
predicted probabilities. We showed that the prediction model for spontaneous pregnancy is able to provide accurate probabilities of ongoing pregnancy over the whole range of subfertile couples.

We feel that this study is important for three reasons. First, the final step in the development of a prediction model for spontaneous pregnancy, i.e. external validation, has been performed. This essential step is often refrained from the development of prediction models, because prospective validation studies are costly and time-consuming (Wasson et al., 1985). Second, this study shows that a prediction model for spontaneous pregnancy can be applied in a general subfertile population, without losing its accuracy, regardless of the setting of the hospital. Third, the information produced by the prediction model for spontaneous pregnancy is accurate. The accurate predictions that we report here justify its use in counselling individual patients. Such an accurate counselling can improve patient care because it enables both doctors and patients to know whether to start fertility treatment. For example, a >40% chance of spontaneous pregnancy in 1 year may justify expectant management, because it is not to be expected that assisted reproduction will increase the already high pregnancy rates in this group. Of course, one should also consider that an expectant management is followed by treatment in case pregnancy does not occur. Pros and cons can be discussed between doctor and patient, and a shared and evidence-based decision can be made.

Originally, the prediction model was developed to predict spontaneous pregnancy leading to live birth. In this study, we validated the model, using ongoing pregnancy as the primary end-point, because at the time of the analysis of this study, follow-up of pregnancies until live birth was not complete for all couples. Because <5% of ongoing pregnancies do not lead to live birth (Yudkin et al., 1987; Feldman, 1992; Hilder et al., 1998), we assume that the prediction model is also well calibrated for the prediction of spontaneous pregnancy leading to live birth.

The performance of the prediction model was assessed by evaluating calibration and discrimination. However, we want to stress that, in the assessment of performance of a prediction model, calibration is more important than discriminative capacity. The latter is more appropriate for evaluating diagnostic tests (Mol et al., 2005). In subfertile couples, it is more important to accurately estimate whether a couple has a high or low chance to conceive than to know exactly which couple will conceive and which will not (Cook et al., 2006).

This study was performed in a multicentre setting to assess whether the prediction model for spontaneous pregnancy is applicable to a general subfertile population. A consequence of the multicentre design was that there was heterogeneity in the performance of tubal assessment. Some clinics completed the fertility work-up by assessing the tubal status with a CAT (Mol et al., 1997; Akande et al., 2003), whereas other clinics always performed either an HSG or a laparoscopy as the final part of the fertility work-up. Therefore, the duration of the fertility work-up in the latter clinics was prolonged, influencing both female age and duration of subfertility. Stratified analyses for these two types of clinics showed comparable baseline characteristics and virtual similar cumulative ongoing pregnancy rates after 1 year (29 and 30%, respectively, data not further shown). We therefore combined these two types of clinics in our analyses.

In this study, we focused on subfertile couples being referred by their general practitioner and excluded couples referred to third care clinics. We assumed that the latter group was referred for fertility treatment and had an assisted reproduction technique indication. Therefore, the use of a prediction model would not have consequences in this group. As a result, there was no variance in the variable ‘referral status’, and subsequently, it only contributed to the baseline probability for all couples.
We also validated the original models of Eimers, Collins and Snick on our data as well as the alternative prediction model that includes the PCT as predictor (Eimers et al., 1994b; Collins et al., 1995; Snick et al., 1997). The model of Eimers was shown to underestimate the pregnancy rates in couples with poor prognoses (0–20%) and overestimate them in couples with good prognoses (≥40%). The calibration was poor. The model of Collins underestimated the pregnancy rates in all couples regardless of their prognoses. The calibration was poor. The model of Snick also showed a poor calibration in a validation study. The prediction model for spontaneous pregnancy, including the PCT, tended to underestimate the chance of conception in couples with a poor prognosis (between 10 and 25%) but overestimated those chances in couples with a good prognosis. Overall calibration was moderate with a slope of 0.58 (95% CI 0.4–0.7, P-value <0.01). All the three original models and the alternative synthesis model with the PCT performed less than the prediction model validated in this study.

Since the introduction of IVF in 1978, its use has increased tremendously. In the USA alone, >58 000 IVF cycles are performed annually (2002). The most important drawback of this phenomenon is high multiple pregnancy rates, (Jain et al., 2002; 2004) with concomitant complications for both mother and child (Mitchell, 2002; Basso and Olsen, 2005). The reduction of the number of transferred embryos has been proposed as a key towards the multiple pregnancy problem (Bhattacharya and Templeton, 2000; Guzick, 2002). Furthermore, IVF is costly. Assuming an IVF cycle to cost $10 000, the average direct cost per baby after IVF is ~$40 000. (Guzick, 2002; Jain et al., 2002) The use of prognostic models with tailored expectant management in couples who are very likely to conceive without assisted reproduction is in our opinion a powerful tool to limit the use of IVF to those subfertile couples with a low likelihood to conceive without IVF.

At this moment, neither clinical guidelines nor current overviews from the American College of Obstetrics and Gynaecology (ACOG) or the Royal College of Obstetrics and Gynaecology (RCOG) (2004) mention the possibility of assessment of the prognosis in the fertility work-up and subsequent counselling according to this prognosis (Bagshawe and Taylor, 2003; National Collaborating Centre for Women’s and Children’s Health, 2004). On the basis of the results from the present study, we believe that the use of prediction models should be discussed in current fertility guidelines.

In clinical practice, the pregnancy prediction model can be used in two ways. The clinician can use a paper score chart (see Supplementary Table II and Figure 4) or a computer version of the prediction model, which calculates the chance to conceive automatically, once the predictive parameters have been entered. This model is now available on the Internet (http://www.freya.nl/probability.php). This format enables gynaecologists all over the world to predict chances on spontaneous pregnancy in a busy practice within a few seconds.

In conclusion, the recently developed synthesis prediction model for spontaneous pregnancy predicts accurately and performs well in a general population. The use of this model in the interpretation of the fertility work-up enables one to identify couples who have a good chance to conceive without treatment. The use of this prediction model may prevent unnecessary treatment. We suggest use of this pregnancy prediction model as an essential tool in counselling subfertile couples.

Acknowledgements
This study was supported by grant 945/12/002 from ZonMW, the Netherlands Organization for Health Research and Development, The Hague, the Netherlands.

Contributors
B.W.J.M., F.V. and P.G.A.H. designed the study. J.W.S. and P.S. promoted it, co-ordinated this cohort study, collected the data and sought ethical approval. F.J.B. and T.H.J.H.M.D. included couples and collected data. J.W.S. did the analysis, under the supervision of B.W.J.M. and M.J.C.E. J.D.F.H. and P.M.M.B. provided statistical advice. All authors helped to prepare the final report. Other contributors in this multicentre study included couples and are mentioned as a part of the CECERM study group.

Supplementary material
Supplementary data are available at http://humrep.oxfordjournals.org/.

References

Submitted on June 18, 2006; resubmitted on August 10, 2006; accepted on August 17, 2006