SEMEN ANALYSIS AND PREDICTION OF NATURAL CONCEPTION

Esther Leushuis
Jan Willem van der Steeg
Pieterneel Steures
Sjoerd Repping
Patrick M.M. Bossuyt
Ben Willem J. Mol
Peter G.A. Hompes
Fulco van der Veen

SUBMITTED
ABSTRACT

STUDY QUESTION
Do two semen analyses predict natural conception better than a single semen analysis and will adding multiple semen parameters to the Hunault prediction model for natural conception improve predictions?

SUMMARY ANSWER
A second semen analysis does not add helpful information for predicting natural conception compared to using the results of a single semen analysis. It did not so for routinely performed second semen analyses, nor for a second semen analysis in selected only. One semen analysis is all that is needed for the prediction of natural conception in the basic fertility workup.

WHAT IS KNOWN ALREADY
A major problem with semen analyses is the large variability of semen parameters within an individual. High quality evidence on how many semen analyses need to be performed during the fertility workup is lacking.

STUDY DESIGN, SIZE, DURATION
Prospective cohort study of 897 consecutive couples presenting with subfertility in two university hospitals in the period 2002 to 2004 in the Netherlands.

PARTICIPANTS/MATERIALS, SETTING, METHODS
We constructed models for three strategies for the prediction of natural conception, using univariable and multivariable cox hazard regression analyses. We evaluated the performance of the three strategies by comparing goodness-of-fit, discrimination and calibration. First, we analysed the semen parameters only. Second, we analysed the semen parameters in addition to the multivariable Hunault prediction model.

MAIN RESULTS AND THE ROLE OF CHANCE
Of the 897 couples 132 (15%) achieved a pregnancy by natural conception. Using the results of a single semen analysis only, the calculated probabilities of natural conception within 12 months ranged from 0.12 to 0.38, with a median of 0.16 [IQR: 0.16 to 0.17]. Using the results of two semen analyses did not lead to a better goodness-of-fit. Discriminative capacity was rather poor, with area under the ROC curve (AUC) of ranging from 0.51 to 0.56. Using the Hosmer-Lemeshow test statistic we found no signs of poor calibration. Using the results of two analyses in combination with the Hunault model did not significantly increase goodness-of-fit compared to using a single semen analysis. The Hunault model with the addition of the semen parameters fit the data significantly better than the Hunault model itself (difference in -2 Log likelihood: 13; 3 df; p=0.002). Using the Hosmer-Lemeshow test statistic we found no signs of poor calibration.
**LIMITATIONS, REASONS FOR CAUTION**
The academic setting possibly explains the relatively low natural conception rates, with only 15% achieving a natural conception within one year. Men with azoospermia were excluded.

**WIDER IMPLICATIONS OF THE FINDINGS**
We believe our study is the first to evaluate the performance of two semen analyses in predicting the chances of natural conception, using data from a large cohort of unselected subfertile couples. We included data from two academic hospitals, located in close proximity, with comparable populations, which limits the generalizability of the prognostic effects but guarantees strong validity of the comparison of strategies.
INTRODUCTION

Semen analysis is the cornerstone of the laboratory evaluation of the male in a subfertile couple. Male subfertility is usually defined by an abnormal result of the semen analysis, although other factors, like impaired sexual function, may play a role even if the semen analysis is normal. The result of the semen analysis in the basic fertility workup contributes to calculations of the probability of natural conception of a subfertile couple. Of all semen parameters sperm motility is the only one in the Hunault model for the prediction of natural conception, next to the prognostic factors female age, duration of subfertility, primary subfertility, and referral status. External validation has shown that the Hunault model has good predictive performance.

A major problem with semen analysis is the large variability of semen parameters within an individual, with additional variability introduced by the examiner and the method of analysis. Because of this variability, it has been suggested to obtain multiple semen analyses. Various guidelines recommend repeating the semen analysis. The NICE guideline and the Dutch (NVOG) guideline for example recommend that a repeat confirmatory test should be offered if the result of the first semen analysis is abnormal. The American guidelines state that the evaluation of the male partner should include at least two semen analyses. This practice was also recommended by the manual of the World Health Organisation in 1999 (WHO), who suggested to supplement two semen analyses with additional semen analyses, if the results were markedly different. In their most recent manual WHO even states that it is impossible to characterize a man’s semen quality from evaluation of a single semen sample, and that it is ‘helpful to examine two or three samples to obtain baseline data’.

The aim of this study was twofold: first to assess the value of repeated semen analyses by analysing to what extent the results of two semen analyses perform better than those of a single semen analysis in predicting natural conception. Second, to assess whether adding the results of two semen analyses, or adding the results of more semen parameters from the first semen analysis to the Hunault prediction model increases its performance in predicting natural conception. We made these comparisons based on data collected from a prospective cohort of subfertile couples in whom two semen analyses had been routinely performed.

MATERIALS AND METHODS

Participants
The data for this study were collected in a prospective cohort study performed in two university hospitals in Amsterdam, the Netherlands: the Vrije Universiteit Medical Center (VUmc) and the Academic Medical Center (AMC). The Institutional Review Board of the Academic Medical Centre had approved the study before its initiation. Local approval was obtained from the Board of Directors from each of the other participating hospitals.
We included consecutive subfertile couples who had not been evaluated previously for subfertility between January 2002 and February 2004. All couples had been referred by their general practitioner and underwent a basic fertility workup. This consisted of a fertility history, two semen analyses, a postcoital test, assessment of ovulation, and assessment of the Fallopian tubes, all according to the prevailing guidelines of the Dutch Society of Obstetrics and Gynaecology.

**Semen analysis**
Semen analysis was performed after three days of sexual abstinence. Semen samples were collected at the hospital or at the participant’s home by masturbation. Semen samples were analyzed within one hour after ejaculation. After liquefaction, semen volume was measured with 0.1 mL accuracy. Sperm concentration was counted and motility was assessed in an appropriate counting chamber at a magnification of x 200. Sperm morphology was registered in one center only (AMC).

The laboratories scored sperm parameters according to the 1999 WHO criteria. Both laboratories participate in the external quality control scheme (EQAS) of the Dutch Foundation for Quality Assessment in Clinical Laboratories (SKML). The laboratory accuracy of sperm concentration, motility and morphology assessment was monitored by the Dutch Foundation for Quality Assessment in Clinical Laboratories (www.skml.nl).

**Follow-up**
After completion of the subfertility workup, we calculated the probability of natural conception within one year. To do so, we used the prediction model developed by Hunault et al.. This model is based on a Cox multivariable regression model developed to predict natural conception leading to live birth within 12 months. The Hunault model includes five prognostic variables: female age, duration of subfertility, female subfertility being primary or secondary, percentage motile spermatozoa of the first semen analysis, and referral status (being referred by general practitioner or gynaecologist). Couples in whom the probability of a natural conception within 12 months was 40% or higher 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. Couples with a probability under 40% were counselled for treatment according to the national fertility guidelines.

Follow-up started at the completion of the infertility workup and ended after 12 months. Primary endpoint in this study was natural conception resulting in an ongoing pregnancy. The first day of that menstrual cycle was considered to mark the end of time until natural conception. Natural 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.

**Data analysis**
Couples in whom the female partner was diagnosed with anovulation or two-sided...
tubal pathology and couples in whom the male partner was diagnosed with azoospermia, defined as the absence of spermatozoa in the ejaculate, were excluded from the analyses, because these couples cannot conceive without treatment.

For each semen parameter, the median and the 5th and 95th percentiles were calculated. We considered a total motile count (TMC) <10*10^6 as abnormal and calculated how many men had an abnormal TMC of their first and second semen sample.

Time to pregnancy was considered censored at the moment treatment had been started 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, and time to pregnancy was not considered as censored at the time of the unsuccessful pregnancy. For all couples lost to follow-up, the general practitioner was sent a questionnaire on the most recent fertility status of the couple.

We defined three strategies to assess whether two semen analyses predict natural conception better than one semen analysis. The three strategies were (1) a single semen analysis, (2) two semen analyses and taking the average as the final result, and (3) a second semen analysis only if the TMC of the first semen analysis is below 10*10^6, again taking the average in that case. The second and third strategy were compared against the first – the reference strategy. For this strategy we only used the results of the first semen analysis in our cohort.

We evaluated the performance of the three strategies by comparing goodness-of-fit, discrimination and calibration in predicting natural conception after the fertility workup. We made two sets of comparisons. The first set was based on the parameters from the semen analysis only. The second set was based on using the semen analysis parameters in addition to the multivariable Hunault prediction model.

**Data analysis – semen analysis only**
We evaluated each strategy by building Cox multivariable regression analyses. The models included semen volume, sperm concentration, and sperm motility.

We measured goodness-of-fit by performing the generalized likelihood ratio test and by calculating Akaike’s Information Criterion (AIC). The strategy with the lowest AIC would be the best fitting parsimonious model.

We additionally evaluated the discriminative power by calculating the area under the curve (AUC) for the receiving operating characteristics curve of each model. A model with perfect discrimination has an AUC of 1, a model with no discrimination an AUC of 0.5.19

We also assessed calibration of the models, using the Hosmer-Lemeshow test statistic. The test assesses whether or not the observed pregnancy rates match pregnancy probabilities in subgroups of the study group. Models for which calculated
probabilities and observed pregnancy rates in subgroups are similar are deemed well calibrated.\textsuperscript{20,21}

\textbf{Data analysis – added value of semen analysis}

To look at the added value of the semen analysis parameters, we constructed a second set of models, one for each of the three strategies described earlier, in which we added the semen analysis parameters sperm motility, semen volume, sperm concentration to the variables in the Hunault model. The Hunault model includes five prognostic variables: female age, duration of subfertility, female subfertility being primary or secondary, sperm motility as determined by the first semen analysis, and referral status.\textsuperscript{7}

For the reference strategy (single semen analysis) we added semen volume and sperm concentration to the model, since sperm motility is already included in the Hunault model. For the two other strategies – using two semen analyses – we added the average of semen volume, sperm motility and sperm concentration of the two semen analyses (strategy 2) or the result from the first or second semen analysis (strategy 3).

We did not re-estimate the weights for the variables from the Hunault model, as this would lead to a loss in power and an increased risk of capitalization on chance: an artificial increase in predictive power which is only due to chance fluctuations in the variables. Instead we used the linear combination (score) of the variables of the Hunault model. In other words, for each couple we calculated the sum of the variables of the Hunault model, each variable weighted by the respective coefficient, as reported by Hunault and colleagues. We then estimated the coefficients for variables in multivariable Cox proportional hazards regression analysis: one coefficient for the Hunault score, and coefficients for the respective semen parameters. We transformed the coefficients into hazard ratios (HR) to facilitate interpretation. For this second set of models we also evaluated goodness-of-fit, discriminative power and calibration of the strategies, using the statistics mentioned earlier. All calculations were performed using SPSS 20 (SPSS Inc., Chicago, IL) program.

\textbf{RESULTS}

\textbf{Participants and semen analyses}

We included data from 963 subfertile couples. Pregnancy outcome data were not available for 66 couples, leaving 897 couples for analysis. Of these, 132 (15\%) achieved a pregnancy by natural conception within one year (Figure 1). The mean male age was 36.0 years (5\textsuperscript{th} to 95\textsuperscript{th} percentile: 28 to 47). The mean age of the female partner was 32.9 (5\textsuperscript{th} to 95\textsuperscript{th} percentile: 25 to 39); the median duration of subfertility was 1.6 years (5\textsuperscript{th} to 95\textsuperscript{th} percentile: 1.0 to 5.0). The results of the two semen analyses and the mean of the two semen analyses are shown in Table I. The variability of two semen analyses can be appreciated from Figure 2, where the results of the second semen analysis are plotted against those of the first analysis, for sperm concentration and for sperm motility.
The TMC was abnormal (< 10^6) in the first semen analysis in 241 men (27%); in 239 (27%) the TMC was abnormal in the second semen analysis. In the 656 men with a normal first semen analysis, the second analysis was abnormal in 76 (12%). The TMC in these men had a median of 4.2*10^6 in the second semen analysis (5th to 95th percentile: 0 to 8.9) versus 22.7 (5th to 95th percentile: 10.7 to 154) in the first. In the 241 men with an abnormal first semen analysis, the second analysis was abnormal in 163 (68%). The TMC in these men had a median of 1.3*10^6 in the second semen analysis (5th to 95th percentile: 0 to 9.2) versus 1.1 (5th to 95th percentile: 0 to 8.9) in the first.

**Performance – semen parameters only**

The results of Cox multivariable regression modelling are presented in Tables II and III. Sperm concentration was in itself a significant predictor of natural conception in univariable analyses, in all strategies. This prognostic effect of sperm concentration was also significant when including other semen parameters in a multivariable analysis. Semen volume and sperm motility had no significant effect on natural conception, regardless of the strategy.

Using the results of a single semen analysis only, the calculated probabilities of natural conception within 12 months ranged from 0.12 to 0.38, with a median of 0.16.
Using the results of two semen analyses did not lead to a better goodness-of-fit compared to strategy 1 (single semen analysis): the values of the AIC for the models that correspond to strategy 2 and 3 were not lower than that for strategy 1 (single analysis), but even higher.

The ROC curves are presented in Figure 3. Discriminative capacity of the first strategy was rather poor, with an area under the ROC curve (AUC) of 0.56 (95% CI: 0.51 to 0.61). The second and third strategy had an AUC of 0.53 (95% CI: 0.48 to 0.58) and 0.51 (95% CI: 0.46 to 0.56) respectively. Differences between the AUC were not significant for strategy 2 versus 1, but were significant for strategy 3 versus 1 (Table III;
<table>
<thead>
<tr>
<th></th>
<th>First semen analysis</th>
<th></th>
<th>Second semen analysis</th>
<th></th>
<th>Mean of two semen analyses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>mean</td>
<td>5th percentile</td>
<td>n</td>
<td>mean</td>
<td>5th percentile</td>
</tr>
<tr>
<td>Volume (mL)</td>
<td>896</td>
<td>3.3</td>
<td>1.0</td>
<td>6.2</td>
<td>897</td>
<td>3.3</td>
</tr>
<tr>
<td>Concentration (10^6/mL) (median)</td>
<td>897</td>
<td>54</td>
<td>1.0</td>
<td>143</td>
<td>897</td>
<td>54</td>
</tr>
<tr>
<td>Motility (grade A %)</td>
<td>897</td>
<td>33</td>
<td>2.0</td>
<td>68</td>
<td>897</td>
<td>34</td>
</tr>
<tr>
<td>Morphology (% normal)^a</td>
<td>658</td>
<td>31</td>
<td>0.0</td>
<td>60</td>
<td>655</td>
<td>32</td>
</tr>
<tr>
<td>TMC^b (10^6) (median)</td>
<td>897</td>
<td>69</td>
<td>0.0</td>
<td>240</td>
<td>897</td>
<td>68</td>
</tr>
</tbody>
</table>

^a according to WHO 1999.
^b TMC: total motile count, calculated as volume*concentration*motility/100.

Table I. Baseline characteristics.
2 versus 1: p = 0.08; 3 versus 1: p = 0.02). Using the Hosmer-Lemeshow test statistic we found no signs of poor calibration, with p-values for the three models of 0.7, 0.1 and 0.5, respectively.

**Performance – added value of semen parameters**

We then evaluated whether addition of the semen analysis—in terms of the three strategies—to the prognosticators from the Hunault model improved prognostic performance. The results for this second set of models are summarized in Table IV. Here also, using the results of two analyses did not significantly increase goodness-of-fit compared to using a single semen analysis: the AIC values were higher, not lower. The model with the results of the first semen analysis added to the Hunault model alone fit the data significantly better than the Hunault model itself (difference in -2 Log likelihood: 13; 3 df; p = 0.002).

Using the results of a single semen analysis in combination with the Hunault prognosticators, the calculated probabilities of natural conception within 12 months ranged from 0.04 to 0.39, with a median of 0.15 [IQR: 0.15 to 0.18]. Discriminatory performance of the three models was better than that of the models based on semen analysis only (Table 3). The AUC for the model based on a single semen analysis was 0.63 (95% CI: 0.60 to 0.68). The second strategy, using the average of two analyses, had an AUC of 0.64 (95% CI: 0.60 to 0.69); the third one of 0.63 (95% CI: 0.58 to 0.68). Differences between these AUC were not significant (Table III; 2 versus 1: p = 0.21; 3 versus 1: p = 0.68). Using the Hosmer-Lemeshow test statistic we found no signs of poor calibration, with p-values for the three models of 0.7, 0.1 and 0.5, respectively.

*Figure 3. ROC Curve.*
Table II. Results of multivariable regression analysis of semen parameters only.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR 95% CI</td>
<td>p</td>
<td>HR 95% CI</td>
</tr>
<tr>
<td>Semen volume</td>
<td>0.95 0.57-1.6 0.85</td>
<td></td>
<td>0.86 0.45-1.7 0.66</td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>0.82 0.72-0.93 0.002</td>
<td></td>
<td>0.83 0.72-0.94 0.005</td>
</tr>
<tr>
<td>Sperm motility</td>
<td>0.94 0.86-1.02 0.14</td>
<td></td>
<td>0.90 0.82-0.99 0.03</td>
</tr>
<tr>
<td>Sperm morphology</td>
<td>0.71 0.48-1.06 0.10</td>
<td></td>
<td>0.49 0.28-0.85 0.01</td>
</tr>
</tbody>
</table>

*Strategy 1: perform a single semen analysis; Strategy 2: Perform two semen analyses. The average of the result from the two analyses is used as the final result. Strategy 3: Conditionally perform a second semen analysis. Initially perform a single analysis, with a second semen analysis if the TMC of the first semen analysis is below 10*10^6. The average of the two results is used as final result in case two analyses are performed.

HR = Hazard Ratio; CI = confidence interval.

Table III. Results of multivariable analysis of the three strategies with semen parameters only.

<table>
<thead>
<tr>
<th></th>
<th>Goodness-of-fit</th>
<th>Discrimination</th>
<th>Calibration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generalized likelihood ratio test (-2log likelihood)</td>
<td>Akaiake's Information Criterion (AIC)</td>
<td>Area under the ROC curve (AUC) (95% CI)</td>
<td>Hasmer-Lemeshow test</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>1608</td>
<td>0.56 0.51-0.61</td>
<td>0.7 0.1</td>
</tr>
<tr>
<td></td>
<td>1602</td>
<td>1610</td>
<td>0.53 0.48-0.58</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>1604</td>
<td>1612</td>
<td>0.63 0.46-0.56</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR 95% CI p</td>
<td>HR 95% CI p</td>
<td>HR 95% CI p</td>
<td>HR 95% CI p</td>
</tr>
<tr>
<td>Semen volume</td>
<td>0.91 0.53-1.54 0.713</td>
<td></td>
<td>0.86 0.44-1.66 0.65</td>
<td></td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>0.82 0.71-0.94 0.006</td>
<td></td>
<td>0.85 0.73-0.99 0.036</td>
<td></td>
</tr>
<tr>
<td>Sperm motility</td>
<td>1 0.91-1.11 0.958</td>
<td></td>
<td>0.96 0.86-1.07 0.467</td>
<td></td>
</tr>
</tbody>
</table>

*Strategy 1: perform a single semen analysis; Strategy 2: Perform two semen analyses. The average of the result from the two analyses is used as the final result. Strategy 3: Conditionally perform a second semen analysis. Initially perform a single analysis, with a second semen analysis if the TMC of the first semen analysis is below 10*10^6. The average of the two results is used as final result in case two analyses are performed.

HR = Hazard Ratio; CI = confidence interval.
Table IV. Results of multivariable analysis with added value of the semen parameters to the Hunault model.

<table>
<thead>
<tr>
<th>Goodness-of-fit</th>
<th>Strategy 1(^a) plus Hunault model(^b)</th>
<th>Strategy 2 plus Hunault model</th>
<th>Strategy 3 plus Hunault model</th>
<th>Hunault model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized likelihood ratio test (-2ll)</td>
<td>1589 *</td>
<td>1594</td>
<td>1598</td>
<td>1602 *</td>
</tr>
<tr>
<td>Akaike’s Information Criterion (AIC)</td>
<td>1597</td>
<td>1604</td>
<td>1608</td>
<td>1606</td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under the ROC curve (AUC) (95% CI)</td>
<td>0.63 0.60-0.68</td>
<td>0.64 0.60-0.69</td>
<td>0.63 0.58-0.68</td>
<td>0.62 0.57-0.70</td>
</tr>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosmer-Lemeshow test</td>
<td>0.140</td>
<td>0.926</td>
<td>0.721</td>
<td>0.633</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lineair predictor (prediction model of Hunault)</th>
<th>Strategy 1(^a) plus Hunault model(^b)</th>
<th>Strategy 2 plus Hunault model</th>
<th>Strategy 3 plus Hunault model</th>
<th>Hunault model</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR 95% CI p</td>
<td>1.701 1.16-2.49 0.006</td>
<td>1.697 1.14-2.53 0.009</td>
<td>1.733 1.16-2.58 0.007</td>
<td>1.756 1.2-2.57 0.004</td>
</tr>
<tr>
<td>Semen volume</td>
<td>1.027 0.92-1.14 0.630</td>
<td>1.056 0.94-1.19 0.365</td>
<td>1.021 0.91-1.14 0.722</td>
<td></td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>1.003 1-1.006 0.060</td>
<td>1.005 1.001-1.009 0.012</td>
<td>1.003 1-1.01 0.098</td>
<td></td>
</tr>
<tr>
<td>Sperm motility</td>
<td></td>
<td>1 0.99-1.01 0.935</td>
<td>0.998 1-1.001 0.774</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Strategy 1: perform a single semen analysis; Strategy 2: Perform two semen analyses. The average of the result from the two analyses is used as the final result. Strategy 3: Conditionally perform a second semen analysis. Initially perform a single analysis, with a second semen analysis if the TMC of the first semen analysis is below 10*10^6. The average of the two results is used as final result in case two analyses are performed.

\(^b\)The Hunault model includes five prognostic variables: female age, duration of subfertility, female subfertility being primary or secondary, percentage motile spermatozoa of the first semen analysis, and referral status (being referred by general practitioner or gynaecologist).

HR = Hazard Ratio; CI = confidence interval.

* difference in -2Loglikelihood: 13, p=0.002.
DISCUSSION

In this cohort study of almost nine hundred subfertile couples we compared the added value of a second semen analysis in terms of the performance gain in predicting natural conception after the basic fertility workup. We showed that using the results of a second semen analysis or a conditional second analysis did not improve predictions of natural conception compared to using the results of a single semen analysis. Adding the first semen analysis to the Hunault model for the prediction of natural conception improved performance significantly compared to using the Hunault model alone, but using an additional second analysis did not further improve performance.

We believe our study is the first to evaluate the performance of two semen analyses in predicting the chances of natural conception based on data from a large cohort of unselected subfertile couples. All semen analyses had been routinely performed as part of the basic fertility workup, and according to the prevailing guidelines. A limitation of this study is that we included couples from only two academic hospitals in close proximity to one another, which may limit the generalizability of the predictions themselves, but guarantees strong validity of the comparison of strategies. The academic setting probably explains the relatively low natural conception rates, with only 15% of couples achieving a natural conception within one year.

Our analyses show that sperm concentration is the most informative of the semen analysis in the prediction of natural conception, but performance does not improve when taking the average of two analyses, or when using a conditional strategy; a second semen analysis only if the TMC of the first semen analysis is below $10^6$, taking the average in that case. Given the variability in semen analysis results, this may come as a surprise. If there is variability, then the average of multiple measurements should be closer to the truth. Yet in this case goodness-of-fit and discrimination did not improve after taking the average of two analyses.

Maybe the variability in semen analysis results is larger than two semen analyses can overcome, and multiple semen analyses, much more than two, are needed to come closer to the ‘truth’, if that would be possible at all. The most recent WHO manual suggests that it is ‘helpful to examine two or three samples to obtain baseline data’ 4, but there is no evidence on the exact number of semen analyses to obtain a true estimate of semen quality. The WHO based its recommendation on five publications.12;13;22-24 Taken together, most of these authors concluded that a minimum of three semen analyses is recommended. However, these studies all ignore the objective of determining the prognostic capacity of semen quality for natural conception. Most of them just evaluate the variability of semen quality of healthy men and none of them take the consequences and possible gains of multiple semen analysis results and time to (natural) conception into account. semen quality and differences between two semen analyses of the male from a subfertile couple might be different from healthy men that do not have a fertility problem. None of these previous studies provided details on how to interpret the results of multiple semen
analyses results; all seem to have used different methods to determine the optimal number of semen analyses for man of a subfertile couple.

Since subfertility is a condition of a couple, male and female factors should be considered in combination when considering the utility of semen analysis in evaluating the subfertile couple. The Hunault model accounts for multiple factors, both male and female, but sperm motility is the only unilateral male prognosticator in this model. In considering this one should keep in mind that this model was constructed as a synthesis model, based on three previous models, which did not have any other semen parameters in common.

We evaluated the performance of the models partially in the group of patients in which it was developed. Internal validation tends to give a too optimistic impression about the quality of the predictions, external validation, in other populations, should therefore follow to confirm our conclusions on the value of sperm concentration, and to evaluate the generalizability or transportability of the extended Hunault model.25;26

In conclusion, this study shows that a second semen analysis does not add helpful information for predicting natural conception, compared to using the results of a single semen analysis. It did not so for routinely performed second semen analyses, nor for a second semen analysis in selected cases only. Yet adding the results of the first semen analysis for semen volume and sperm concentration to the Hunault model, which already includes sperm motility, may be able to improve predictions of natural conception significantly, compared to using the Hunault model without these additional semen parameters.
REFERENCES


