Semen analysis and prediction of natural conception

Esther Leushuis1,2,*, Jan Willem van der Steeg3, Pieternel Steures4, Sjoerd Repping2, Patrick M.M. Bossuyt5, Ben Willem J. Mol2, Peter G.A. Hompes1, and Fulco van der Veen2

1Department of Obstetrics and Gynaecology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands 2Center for Reproductive Medicine, Academic Medical Center, Amsterdam, The Netherlands 3Department of Obstetrics and Gynaecology, Jeroen Bosch Ziekenhuis, ’s Hertogenbosch, The Netherlands 4Department of Obstetrics and Gynaecology, St. Elisabeth Ziekenhuis, Tilburg 5Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands

*Correspondence address. VU Medisch Centrum, Department of Obstetrics and Gynaecology, PK 0 Z 116.1, Postbus 7057, 1007 MB Amsterdam, The Netherlands. Tel: +31-20-444-00-70; Fax: +31-20-444-00-45; E-mail: e_leushuis@hotmail.com

Submitted on April 27, 2013; resubmitted on December 21, 2013; accepted on March 5, 2014

STUDY QUESTION: Do two semen analyses predict natural conception better than a single semen analysis and will adding the results of repeated semen analyses to a prediction model for natural pregnancy improve predictions?

SUMMARY ANSWER: A second semen analysis does not add helpful information for predicting natural conception compared with using the results of a single semen analysis and addition of the second analysis to a prediction model for natural conception did not improve predictions.

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

STUDY DESIGN, SIZE, DURATION: We conducted a prospective cohort study of 897 consecutive couples presenting with subfertility in two university hospitals in the period 2002–2004 in the Netherlands.

PARTICIPANTS/MATERIALS, SETTING, AND METHODS: The laboratories scored sperm parameters according to the 1999 WHO criteria. Sperm concentration was counted and motility was assessed in a Makler counting chamber at a magnification of ×200. All assessments were performed by trained laboratory technicians. Follow-up started at the completion of the infertility workup and ended after 12 months. Primary end-point was natural conception resulting in an ongoing pregnancy. 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. Secondly, we analysed the semen parameters in addition to the 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 across the study population ranged from 0.12 to 0.38, with a median of 0.16 (IQR: 0.16–0.17). Using the results of two semen analyses did not lead to a better goodness-of-fit. Discriminative capacity was rather poor, with an area under the ROC curve (AUC) 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 semen analyses in combination with the Hunault model did not significantly increase goodness-of-fit compared with using a single semen analysis. The Hunault model with the addition of the semen parameters fitted 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 1 year. Men with azoospermia were excluded.

WIDER IMPLICATIONS OF THE FINDINGS: Performing more than one semen analysis will not increase the prognostic power of the test in clinical practice. Adding the first semen analysis to the Hunault model for the prediction of natural conception improved performance significantly compared with using the Hunault model alone. External validation, in other populations, should follow to confirm our conclusions, and to evaluate the generalizability or transportability of the extended Hunault model.
**Introduction**

Semen analysis is the cornerstone of the laboratory evaluation of the male in a subfertile couple (de Kretser, 1997; The Male Infertility Best Practice Policy Committee of the American Urological Association and The Practice Committee of the American Society for Reproductive Medicine, 2006). 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 semen analysis in the basic fertility workup has multiple aims. On the one hand, semen analysis contributes to calculations of the probability of natural conception of a subfertile couple [Dutch Society of Obstetrics and Gynaecology (NVOG), 2004; The Male Infertility Best Practice Policy Committee of the American Urological Association and The Practice Committee of the American Society for Reproductive Medicine, 2006; World Health Organization, 2010; Dutch Society of Obstetrics and Gynaecology (NVOG), 2010; Dutch Society of Obstetrics and Gynaecology (NVOG) and Dutch College of General Practitioners (NHG), 2010]. Semen analysis also guides decisions about assisted reproduction techniques and it detects causes, like hypogonadotropic hypogonadism or cryptorchidism, that may call for treatment of man in order to improve his health (irrespective of success of any treatment of subfertility with assisted reproduction techniques) [Dutch Society of Obstetrics and Gynaecology (NVOG) and Dutch College of General Practitioners (NHG), 2010]. The semen analysis alone provides a poor prediction of the chance that a couple will conceive (Bonde et al., 1998; Guzick et al., 2001). A validated prediction model for natural conception (Hunault model) only includes sperm motility of all semen parameters, next to the prognostic factors female age, duration of subfertility, primary subfertility, and referral status. A fuller explanation of the Hunault model is given in the Materials and Methods (Hunault et al., 2004). External validation has shown that the Hunault model has good predictive performance (van der Steeg et al., 2007).

A major problem with semen analysis is the large variability of semen parameters between different ejaculates from the same individual. Moreover, additional variability is introduced by differences between examiners and between methods of analysis (Auger et al., 2000; Rheimrev et al., 2001; Leushuis et al., 2010; World Health Organization, 2010). Because of this variability, it has been suggested that multiple semen analyses should be done (Castilla et al., 2006; Poland et al., 1985). 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 [Dutch Society of Obstetrics and Gynaecology (NVOG), 1999; National Collaborating Centre for Women’s and Children’s Health, 2004; National Collaborating Centre for Women’s and Children’s Health, 2013]. The American guidelines state that the evaluation of the male partner should include at least two semen analyses (The Male Infertility Best Practice Policy Committee of the American Urological Association and The Practice Committee of the American Society for Reproductive Medicine, 2006). This practice was also recommended by the manual of the World Health Organization (WHO) in 1999, which suggests supplementing two semen analyses with additional semen analyses, if the results were markedly different (World Health Organization, 1999). In their most recent manual WHO even state 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’ (World Health Organization, 2010).

The aim of this study was 2-fold: 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. Secondly, 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). Under the legal requirements for clinical research in the Netherlands, this study was exempt from medical ethics and institutional review board approval.

We included consecutive subfertile couples who had not been evaluated previously for subfertility between January 2002 and February 2004 (van der Steeg et al., 2007). 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 (Dutch Society of Obstetrics and Gynaecology (NVOG), 2004) (Table I).

**Semen analysis**

The laboratories scored sperm parameters according to the 1999 WHO criteria (World Health Organization, 2000). Men were advised to maintain 3 days of sexual abstinence prior to semen analysis. The interval of the time of abstinence was recorded when the sample was handed in. Semen samples were collected at the hospital or at home by masturbation directly into a 50-ml polyethylene jar. Semen samples were analysed within one hour after ejaculation. After liquefaction, semen volume was measured in a graded tube with a 0.1-ml accuracy. Sperm concentration was counted and motility was assessed in a Makler counting chamber at a magnification of x 200. All assessments were performed by trained laboratory technicians. Sperm morphology was measured in one laboratory only. 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

---

**STUDY FUNDING/COMPETING INTEREST(S):** No external funding was involved in this study. None of the authors has any conflict of interest to declare.

**Key words:** semen analysis / pregnancy / prediction / subfertility
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 (i) a single semen analysis, (ii) two semen analyses and taking the average as the final result and (iii) a second semen analysis only if the TMC of the first semen analysis is \(< 10^{-5}\) again taking the average in that case. The second and third strategies 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. All parameters were treated as continuous variables.

We compared the odds ratio (OR) and 95% confidence interval (CI) of the chance of conception based on the TMC of normal semen analyses versus two abnormal semen analyses.

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 (Leushuis et al., 2009).

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 (Coppus et al., 2009; Hosmer, 2000).

### 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 and 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 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

---

**Table I** Criteria for inclusion and exclusion of subfertile couples.

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subfertile couples not evaluated previously for subfertility</td>
<td>Causal pathology:</td>
</tr>
<tr>
<td>Referred by the general practitioner</td>
<td>Ovulation disorder</td>
</tr>
<tr>
<td>Complete basis fertility workup consisting of:</td>
<td>Two-sided tubal pathology</td>
</tr>
<tr>
<td>Fertility history</td>
<td>Azoospermia</td>
</tr>
<tr>
<td>Two semen analyses</td>
<td>Missing semen parameters</td>
</tr>
<tr>
<td>Postcoital test</td>
<td>Primary outcome missing</td>
</tr>
<tr>
<td>Assessment of ovulation</td>
<td>Duration of follow-up missing</td>
</tr>
<tr>
<td>Assessment of fallopian tubes</td>
<td></td>
</tr>
</tbody>
</table>

---

Follow-up

After completion of the subfertility workup, we calculated the probability of natural conception within 1 year. To do so, we used the prediction model developed by Hunault et al. (2004). 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 (Dutch Society of Obstetrics and Gynaecology (NVOG), 2004).

Follow-up started at the completion of the infertility workup and ended after 12 months. Primary end-point 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 an ongoing pregnancy was defined as the presence of fetal cardiac activity at
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, USA) program.

Results

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 1 year (Fig. 1). The mean male age was 36.0 years (5–95th percentile: 28–47). The mean age of the female partner was 32.9 (5–95th percentile: 25–39); the median duration of subfertility was 1.6 years (5–95th percentile: 1.0–5.0). The results of the two semen analyses and the mean of the two semen analyses are shown in Table II. The variability of two semen analyses can be appreciated from Fig. 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 (5–95th percentile: 0–8.9) versus 22.7 (5–95th percentile: 10.7–154) 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 (5–95th percentile: 0–9.2) versus 1.1 (5–95th percentile: 0–8.9) in the first.

Performance: semen parameters only

The results of Cox multivariable regression modelling are presented in Tables III and Supplementary data, Table SI. 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. The OR of the chance of conception based on the TMC of two normal semen analyses versus two abnormal semen analyses was 2.3 (95% CI: 1.4–3.7).

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: 0.16–0.17). Using the results of two semen analyses did not lead to a better goodness-of-fit compared with
strategy 1 (single semen analysis): the values of the Akaike’s Information Criterion (AIC) for the models that correspond to strategies 2 and 3 were not lower than that for strategy 1 (single analysis), but even higher. The strategy with the lowest AIC would be the best fitting parsimonious model.

The ROC curves are presented in Fig. 3. Discriminative capacity of the first strategy was rather poor, with an area under the ROC curve (AUC) of 0.56 (95% CI: 0.51–0.61). The second and third strategies had an AUC of 0.53 (95% CI: 0.48–0.58) and 0.51 (95% CI: 0.46–0.56), respectively. Differences between the AUC were not significant for strategy 2 versus 1, but were significant for strategy 3 versus 1 (Supplementary data, Table SI: 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 III. Here also, using the results of two analyses did not significantly increase goodness-of-fit compared with 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: 0.15–0.18). Discriminatory performance of the three models was better than that of the models based on semen analysis only (Supplementary data, Table SII). The AUC for the model based on a single semen analysis was 0.63 (95% CI: 0.60–0.68). The second strategy, using the average of two analyses, had an AUC of 0.64 (95% CI: 0.60–0.69); the third one of 0.63 (95% CI: 0.58–0.68). Differences between these AUC were not significant (Supplementary data, Table SII: 2 versus 1: \(P = 0.21\); 3 versus 1: \(P = 0.68\)). Using the Hosmer–Lemeshow test
We showed that using the results of a second semen analysis or a conditional strategy; a second semen analysis only if the TMC of the first semen analysis is below \(10 \times 10^6\). The average of the two results is used as the final result in case two analyses are performed.

Table III Results of multivariable Cox regression analysis of semen parameters only for the prediction of natural conception.

<table>
<thead>
<tr>
<th>Semen volume (ml)</th>
<th>Strategy 1 HR 95% CI P</th>
<th>Strategy 2 HR 95% CI P</th>
<th>Strategy 3 HR 95% CI P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.95 0.57–1.6 0.85</td>
<td>0.86 0.45–1.7 0.66</td>
<td>0.98 0.54–1.78 0.95</td>
</tr>
<tr>
<td>Sperm concentration ((10^6 / \text{ml}))</td>
<td>0.82 0.72–0.93 0.002</td>
<td>0.83 0.72–0.94 0.005</td>
<td>0.83 0.73–0.95 0.007</td>
</tr>
<tr>
<td>Sperm motility (Grade A %)</td>
<td>0.94 0.86–1.02 0.14</td>
<td>0.90 0.82–0.99 0.03</td>
<td>0.94 0.86–1.02 0.13</td>
</tr>
<tr>
<td>Sperm morphology (% normal)</td>
<td>0.71 0.48–1.06 0.10</td>
<td>0.49 0.28–0.85 0.01</td>
<td>0.63 0.39–1.00 0.05</td>
</tr>
</tbody>
</table>

HR, hazard ratio; CI, confidence interval.

aStrategy 1: perform a single semen analysis; Strategy 2: perform two semen analyses. The average of the results 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 \(10 \times 10^6\). The average of the two results is used as the final result in case two analyses are performed.

bAccording to World Health Organization (WHO) 1999.

Discussion

In this cohort study of almost 900 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 with 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 with 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. One study of 113 fertile men and 109 subfertile men that was designed to identify the best predictors of fertility using selected parameters of semen analysis by logistic regression modelling revealed that the best prediction of spontaneous conception was obtained using 12 semen parameters (Jedrzejczak et al., 2008). This is in contrast to our study that specifically evaluated the added value of a second semen analysis and that compared the performance of models using a selected set of semen parameters and included the prognostic factors of a validated model for the prediction of natural pregnancy.

A limitation of our 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 1 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 \(10 \times 10^6\), 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. Whether there is an optimal number of semen analyses to reflect the ‘true semen result’ and what this number would be is a complex issue, which could be the subject of future studies.

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’ (World Health Organization, 2010), but there is no evidence on the exact number of

Figure 3 Receptor operator characteristic curves to illustrate the predictive power of the first semen analysis (Strategy 1), the mean of the first and second semen analysis (Strategy 2) and the mean of the first and second semen analysis where the first was abnormal otherwise the first semen analysis only (Strategy 3).

In this cohort study of almost 900 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 with 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 with 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. One study of 113 fertile men and 109 subfertile men that was designed to identify the best predictors of fertility using selected parameters of semen analysis by logistic regression modelling revealed that the best prediction of spontaneous conception was obtained using 12 semen parameters (Jedrzejczak et al., 2008). This is in contrast to our study that specifically evaluated the added value of a second semen analysis and that compared the performance of models using a selected set of semen parameters and included the prognostic factors of a validated model for the prediction of natural pregnancy.

A limitation of our 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 1 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 \(10 \times 10^6\), 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. Whether there is an optimal number of semen analyses to reflect the ‘true semen result’ and what this number would be is a complex issue, which could be the subject of future studies.

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’ (World Health Organization, 2010), 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 (Poland et al., 1985; Berman et al., 1996; Carlsen et al., 2004; Castilla et al., 2006; Keel, 2006). 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 analyses 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 (Altman and Royston, 2000; Harrell et al., 1996).

In conclusion, this study shows that a second semen analysis does not add helpful information for predicting natural conception, compared with 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 with using the Hunault model without these additional semen parameters.

Supplementary data

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

Authors’ roles

B.W.J.M., F.V. and P.G.A.H. designed the study. E.L. contributed to the design of the study, analysed the data and drafted the manuscript. J.W.S. and P.S. co-ordinated the cohort study, collected the data and sought ethical approval. E.L. performed the analysis under the supervision of P.M.M.B. All authors contributed to the revisions of the manuscript. All authors were involved in the approval of the manuscript.

Funding

No external funding was involved in this study.

Conflict of interest

None of the authors has any conflict of interest to declare.

References

Dutch Society of Obstetrics and Gynaecology (NVOG) and Dutch College of General Practitioners (NHG). National Network Guideline on Subfertility. 2010.


