SUMMARY AND IMPLICATIONS
Subfertility, defined as failure to conceive after at least twelve months of unprotected intercourse affects 10% to 15% of all couples trying to conceive. In 25% of the subfertile couples a solitary male factor, in 25% a female factor, and in 25% a combination of male and female factors exists, while in 25% of the couples no cause can be found. In these couples the subfertility is designated as unexplained. Female causes of subfertility result from disorders of ovulation or pathology of the fallopian tubes; other causes include endometriosis, cervical defects and uterine abnormalities. If male subfertility is present, it is almost always defined by an abnormal finding in the semen analysis, although other male factors like impaired sexual function may play a role even if the semen analysis is normal.

Male subfertility can be due to a variety of conditions. Fertility can be reduced as a result of congenital or acquired urogenital abnormalities, urogenital tract infections, increased scrotal temperature, endocrine disturbances, genetic abnormalities or immunological factors. In 30-40% of cases, no factor is found, a condition called idiopathic male subfertility. Some of the conditions are identifiable and reversible, such as ductal obstruction, hypogonadotrophic hypogonadism and varicocele. Other conditions are identifiable but not reversible such as bilateral testicular atrophy secondary to viral orchitis. The main purpose of the evaluation of the male for subfertility is to identify a reversible condition when present. Identification and treatment of reversible conditions may improve male’s fertility and allow for natural conception. Detection of conditions for which there is no treatment will spare couples the distress of attempting ineffective therapies. Detection of certain genetic causes of male subfertility allows couples to be informed about the potential to transmit genetic abnormalities that may affect the health of offspring.

The basic fertility workup is the initial investigation that is performed in subfertile couples to evaluate possible male and female disorders. This workup includes fertility history, semen analysis, MAR test, postcoital test (PCT), assessment of ovulation, and assessment of the Fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and Gynaecology.

Prediction of (natural) conception
After completion of the fertility workup it is essential to distinguish subfertile couples in whom the prognosis of natural conception is poor and fertility treatment is mandatory from subfertile couples who still have a good prognosis to conceive naturally. In 2004 Hunault et al. published a key model for the prediction of natural conception. Several prediction models on natural and treatment-dependent conception were published before and after this publication. Careful evaluation is needed before implementing these models in clinical practice. Poor-quality models could have a negative effect on decision making by introducing the illusion of objective improvement over clinical judgement. In this thesis we evaluated the available prediction models in reproductive medicine by reviewing the literature and appraising the quality of the models. We identified 36 papers reporting on 29 prediction models. Out of nine models for the prediction of natural conception, three for the prediction
of pregnancy after IUI and 17 for the prediction of pregnancy after IVF, only three models were assessed to be of good performance: one on natural pregnancy (Hunault model)\(^9\), one on IUI (Steures model)\(^10\) and one on IVF (Templeton model)\(^11\). These models can be considered as a guiding tool in making decisions for fertility treatment in couples (Chapter 2).

**The male factor in the prediction of natural conception**

In our review, we explored the role of the male factor in the available prediction models for (natural) conception. Two out of nine of these models did not report any male prognosticators.\(^12\) The seven other models differed widely in report on male prognosticators. Decreasing values of the following prognosticators were significantly associated with a lower probability of natural conception: male age, sperm motility, normal sperm morphology, sperm concentration, hypo-osmotic swelling (HOS) test, presence of urethritis in history and fertility problems in man’s family (Chapter 2).\(^12\) However, the majority of the seven models that incorporated these male prognosticators was not evaluated in an external validation. Good performance of a prediction model in external validation is a minimal requirement to be eligible for use in clinical practice.

These results exposed the lack of knowledge on male prognosticators as the Achilles’ heel in predicting natural conception, despite extensive research. With the resurgence of interest for prognosis and improved opportunities for the development of prediction models, we aimed to increase evidence on the male prognosticators for natural conception in this thesis. We focused on three most important prognosticators that are associated with the male factor: the semen analysis, the MAR test and the PCT.

**Semen analysis**

Semen analysis may be considered as the cornerstone of the laboratory of the male in a subfertile couple.\(^3\) According to the Dutch national guideline for the basic fertility workup, the semen analysis (sperm motility) is one of the prognostic factors that is used in the model for the prediction of natural conception.\(^2\) To critically appraise the role of the male factor in the prediction of natural conception, it is important to determine the precise degree of variability that is represented by the reproducibility and reliability of semen analyses. Until now, the evidence on the precise degree of variability is not well established in male partners of subfertile couples. We analyzed the data of 5,240 male partners of a subfertile couple and found that the reliability in terms of the within-subjects coefficient of variation (CV\(_w\)) of all semen parameters was high and ranged from 28% to 34%. The reproducibility in terms of the intraclass correlation coefficients (ICCs) of these semen parameters were moderate to high (volume: 0.70; concentration: 0.89; motility: 0.58; morphology: 0.60; total motile count: 0.73). In summary, this study affirmed the presumed large within-subject variability and the limited reproducibility of semen analyses in subfertile men (Chapter 3).

With this new evidence on the large variation of semen analysis, we hypothesized that semen analyses need to be repeated to reflect the true value of semen parameters.\(^25\) Various guidelines indeed recommend repeating the semen analysis.\(^3\) The NICE
guideline and the Dutch (NVOG) guideline 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 (WHO) in 1999, who suggested to supplement two semen analyses with additional semen analyses, if the results were markedly different. In their most recent manual they 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’.

We performed a study, 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—which already includes sperm motility of the first semen analysis—increases performance. 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. We made these comparisons in a prospective cohort of subfertile couples in whom two semen analyses had been routinely performed.

Data were collected from 897 couples in two academic hospitals. 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 10^6, again taking the average in that case. For strategy 1 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 in three steps. The first strategy—a single analysis—was used as reference strategy. We made two sets of analyses. The first set was based on the semen analysis only. The second set of models combined the three models of the strategies—that included the parameters semen volume, sperm concentration, and sperm motility—with the Hunault model. In this second set of models, the first model consisted of the Hunault model complemented with the semen volume and sperm concentration (strategy 1 without the sperm motility of the first semen analysis which is already included in the Hunault model); the second model consisted of the Hunault model complemented with the average of semen volume, sperm motility and sperm concentration of the two semen analyses (strategy 2); and the third model consisted of the Hunault model complemented with the first or second semen analysis according to strategy 3.

Using the results of two semen analyses, in strategy 2 or 3, 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. 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 respectively 0.53 (95% CI: 0.48 to 0.58) and 0.51 (95% CI: 0.46 to 0.56). Differences between the AUC were not significant for strategy 2 versus 1, but were significant for strategy 3 versus 1. 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.

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. Here also, using the results of two analyses, as in strategy 2 or 3, 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 between the model of the first semen analysis added to the Hunault model and the Hunault model alone: 13; 3 df; p=0.002). Discriminatory performance of these three models of the second set was better than that of the models based on semen analysis only. The AUC for the first strategy was 0.63 (95% CI: 0.60 to 0.68). The second strategy 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 (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.

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, nor for addition of these two strategies to the Hunault prediction model for natural conception. However, adding the result of the first semen analysis for semen volume and sperm concentration to the Hunault model, that already includes sperm motility, did improve prediction of natural conception significantly compared to the Hunault model alone (Chapter 4).

Results of the semen analysis can vary considerably between laboratories. As a consequence, it is difficult for doctors to interpret and compare the results of semen analyses from different laboratories and this hampers the value of the semen analysis in daily practice. Standardization of semen analysis results might improve reproducibility. We assessed variability in semen analyses between laboratories and determined whether standardization, by applying transformations using Z-scores and regression statistics, was able to normalize results and improve the interpretability and usability of semen analyses.

Semen samples were circulated on sperm concentration and sperm morphology to
participating laboratories in an External Quality Assessment Scheme (the circulation cohort) and we computed transformations. We calculated between-laboratory coefficients of variation (CVb), defined as the ratio of the standard deviation to the mean which expresses the between-laboratory standard deviation relative to the laboratories’ true score for each parameter. CVb values close to zero indicate better reproducibility and less variability between laboratories. The mean CVb was 7% for sperm concentration (range 3% to 13%) and 32% for sperm morphology (range 18% to 51%) in the circulation cohort of the semen samples. We validated the two computed transformations in the circulation cohort by testing if systematic differences were reduced after transformations. Z-score transformation in the circulation cohort was only valid for sperm morphology and regression statistic transformation was valid for both semen parameters. We applied the validated transformations to the results of the semen analysis of a prospective study cohort. This cohort consisted of men from subfertile couples from a prospective cohort study, collected between January 2002 and February 2004 from 20 laboratories, in whom a semen analysis was performed by one of the laboratories of the circulation cohort. This study cohort included 2,804 men with data on sperm concentration and sperm morphology. The baseline differences between the laboratories of the study cohort were statistically significant for all semen parameters (p<0.0001). Standardization using either transformation did not reduce the differences in semen analysis results between the laboratories (p<0.0001). A potential limitation could be that the random error overshadowed the presence of an effect. Although all measurements are prone to random error, the source of random error in the variability of the semen analysis itself and between laboratories is difficult to assess. Another explanation for the inability of both transformation methods to correct for systematic differences might be differences in the populations of the laboratories. In conclusion, a large between-laboratory variation exists for sperm morphology and small, but still significant between-laboratory variation for sperm concentration. Standardization with a transformation using Z-score or regressions statistics did not reduce such variation (Chapter 5).

**MAR-test in semen**

The mixed antiglobulin reaction test in semen (direct MAR test) was also assumed to affect chances of natural conception. Although the World Health Organisation (WHO) recommends testing for IgG by means of the MAR test as a routine screening method for antisperm antibodies, and if tested positive for IgG, followed by an IgA test, the clinical significance of IgG ASA is still not clear and not all guidelines for subfertility recommend this test in the basis fertility workup. In view of this uncertainty, we performed a study to assess the capacity of immunoglobulin G (IgG) antisperm antibodies in the direct MAR test to predict natural ongoing pregnancy in a large prospective cohort of subfertile couples. We included 1,794 couples, of which 283 (16%) had a natural ongoing pregnancy within one year. When a threshold 50% was used for an abnormal test result, the MAR test was positive in 3% of the couples. In the multivariable analysis, a positive MAR test (≥50%) had no contribution in the prediction of natural pregnancy (HR 0.99, 95% CI 0.40 to 2.4). In conclusion, this large cohort study showed that the MAR test is not able to predict the probability of natural conception.
conception. The routine performance of the MAR test in the basic fertility workup for identification of couples with low natural pregnancy chances is not justified (Chapter 6).

**Postcoital test**
The most disputed test predicting natural conception is the postcoital test (PCT). The PCT is traditionally used to diagnose cervical factor subfertility. In this test the number of forward-moving spermatozoon in the cervical mucus sample that is collected within 8 to 16 hours after intercourse is assessed. The PCT is judged normal if at least one progressively motile spermatozoon is seen in one of five high-power fields at ×400 magnification. The PCT is judged abnormal if non-motile or no spermatozoa are seen in one of the five high-power fields at ×400 magnification. This test has been the subject to debate for more than 15 years due to conflicting data.21;33;34 At present, data on the prognostic value of the PCT for natural conception in subfertile couples are inconclusive. We performed a study to evaluate the capacity of the postcoital test (PCT) to predict natural conception in a large cohort study of 3,021 subfertile couples and to assess whether the PCT is an essential part of the basic fertility workup to predict natural conception. We constructed a ‘PCT model’ that consisted of the PCT result as a potential prognosticator augmented with the established prognosticators of the reference (Hunault) model. The adjusted HR for an abnormal PCT was 0.76 (95% confidence interval [CI]: 0.62 to 0.94). Adding the PCT result to the reference model (the PCT model) did not improve goodness-of-fit. Discrimination was equally poor for the PCT model and the reference model. The calibration plots of both models showed comparably good calibration. The net reclassification improvement of the predictions of the PCT model compared with the reference model was -1.1%, meaning that the PCT model showed a minor net reduction in the correct classification of couples who did and did not achieve a pregnancy compared with the reference model. In summary, this study demonstrated that the PCT has prognostic value but does not add substantially to a prognostic model for natural conception. The routine performance of the PCT in the basic fertility workup for identification of couples with low natural pregnancy chances is thus not justified (Chapter 7).

**CONSEQUENCES FOR CLINICAL PRACTICE AND RECOMMENDATIONS FOR FUTURE RESEARCH**

So currently, there are three models with good predictive performance. One prediction model for natural pregnancy 12, one for IUI 13 and one for IVF 14. The Hunault prediction model is the best validated model for the prediction of natural conception. This model including female age, duration of subfertility, primary subfertility, sperm motility, and referral status has good predictive performance. The model can be used reliably as a guide for making decisions about fertility treatment, in patients similar to the development population. The Steures prediction model is the best validated model for pregnancy after IUI and includes female age, duration of subfertility, cervical factor, male factor, tubal pathology, endometriosis, uterine anomaly, mild ovarian
stimulation and an increasing number of cycles. The Templeton prediction model is the best prediction model for pregnancy after IVF. It includes female age, duration of subfertility, tubal reasons for subfertility, live birth after IVF, live birth that was not a result of IVF, a previous pregnancy after IVF that did not result in a live birth, and a previous pregnancy not after IVF that did not result in a live birth.

In this thesis we focused on prediction of natural conception. The addition of the PCT result to the prediction model for natural conception (Hunault model) did not improve its performance. We showed that the PCT does have prognostic value, but does not add substantially to the known prognosticators from this model. Performance of the PCT in the basic fertility workup to assess probabilities of natural conception seems therefore not justified.

The predictive capacity of the MAR test for natural conception was also limited compared to the well-known prognosticators from the Hunault model. Its routine use in the basic fertility workup for identification of couples with low natural conception probabilities seems therefore not justified.

Although every doctor in reproductive medicine intuitively knows semen is variable, still, the evidence on the precise degree of variability was not well established in male partners of subfertile couples. We showed that the within-subject variability of the semen analysis in subfertile men is large and the reproducibility of semen analyses in subfertile men is limited.

We also confirmed systematic differences in semen analyses between laboratories and standardization with Z-score or regression coefficients, could not reduces these systematic differences. The repetition of the semen analysis when consulting another hospital i.e. laboratory therefore seems inevitable and will remain a burden for the male partner of the subfertile couple.

In addition to this and/or to diminish the amount of investigations that do not provide extra information for the prognosis of the subfertile couple, we evaluated whether the semen analysis routinely had to be repeated to evaluate the couple’s probability of natural conception in the basic fertility workup.

The current network guideline for subfertility in the Netherlands advises to repeat the semen analysis if the result is abnormal according to WHO manual. This recommendation is based on a re-evaluation of men from a retrospective chart review that had two out of three abnormal semen analyses according to the WHO criteria of 1987. This chart review in 209 aimed to evaluate the ‘diagnostic predictability’ of one to three semen analyses for the identification of ‘male infertility’. As the authors state, the results of their study are subjected to bias and the numerous flaws in the execution of this chart review (study design, biased comparisons, omission of important study details and definitions, no substantiation of methods used, not related to conception or pregnancy etcetera) degrade the validity of this study. The guideline also emphasises that the primary aim of the performance of the semen analysis is to predict the probability of natural conception, which is dependent on four primary prognosticators in addition to the parameter of the semen analysis, sperm
motility: female age, duration of subfertility, primary subfertility, and referral status. The guideline promotes interpretation of the semen analysis integrated with the other – non male – prognosticators in the prediction model for natural pregnancy (Hunault model).

Our study was designed to address the issue of the repetition of the semen analysis in the basic fertility workup for the prediction of natural conception specifically and provides the evidence to endorse this recommendation. We concluded that routine repetition of the semen analysis seems to have no additional value for the prediction of natural conception in the basic fertility workup.

In our view, three issues in the setting of (male) prognostic research in the field of prediction of natural conception deserve priority to be clarified further. All should eventually lead to optimal evaluation of the male partner of a subfertile couple. First, it has not been proven in a randomised clinical trial or in a clinical setting after implementation of the prediction models, whether prediction models actually contribute to improving the care for individual couples (impact analysis). We should encourage further implementation as well as a more extensive documentation and specifically study its impact on care after implementation to assess actual improvement of care by prediction models. This can be facilitated by the Marker-by-Treatment-Interaction design to evaluate the impact of the prediction model. This design implies that the prediction model’s calculated prognosis (‘marker’) splits the population into groups, in which the efficacy of a particular treatment will differ. It can be viewed as a classical randomized clinical trial with upfront stratification for the prediction model’s calculated prognosis.36

Second, if at all possible, we could design and facilitate a trial that would allow construction of one general prediction model for the subfertile couple to predict (natural or assisted) conception. After completion of the fertility workup it is essential to distinguish subfertile couples in whom prognosis of natural conception is poor and fertility treatment mandatory from subfertile couples who still have a good prognosis to conceive naturally.11 One integrated prediction model not only for natural conception, but also for pregnancy after IUI and IVF/ICSI, can support the clinician even more to identify which individuals will benefit from the given treatment (including expectant management), with the goal of maximizing good outcomes and minimizing side effects, treatment burden, and medical costs. From a patient’s perspective, presentation of these probabilities in a patient friendly format, i.e. by developing adequate patient information material, could encourage a couple to refrain from assisted reproduction techniques (ART) with higher costs, side effects and potentially lower chances of conception.37

The current evidence concerning the best treatment option for couples with unexplained and male subfertility is inconclusive. Most studies that have evaluated the effectiveness of treatment options, such as expectant management (EM), IUI, with or without controlled ovarian stimulation (COS), and IVF, have not taken the couples’ prognosis into account. It is very likely that the individual prognosis of the couple influences the effect of treatment. Individual patient data analyses (IPD) allows to
take these prognostic factors into account, and to evaluate their effect on treatment outcome. Such an IPD is currently performed by van den Boogaard et al. and aims to use anonymised data from relevant published trials to perform an IPD meta-analysis, evaluating the effect of couples’ prognosis on the effectiveness of EM, IUI, with or without COS, and IVF. The probability of natural conception is lower when the semen analysis is abnormal and male subfertility is diagnosed, as the prediction model of Hunault illustrates. Since there are no treatments to improve semen quality, current treatment of male subfertility is based on optimizing the chances of fertilization by delivering motile sperm close to (IUI), very close to (IVF), or even inside (ICSI) the oocyte. The first large scale randomized multicentre trial on the effectiveness of treatments for male subfertility will be initiated this year (MASTER study). The primary aim of this project is to assess the cost-effectiveness of therapies for male subfertility and is its design is split up into three parallel RCTs: IUI versus expectant management in mild male subfertility, IVF versus IUI in moderate male subfertility and ICSI versus IVF in severe male subfertility.

Third, the elusiveness of a man’s true semen analysis result and its relatively unknown prognostic value to assess his fertility potential has been affirmed in this thesis. An opulence of studies have aimed to assess a subfertile man’s true semen status, but variability in person and place seems to restrict the feasibility of solid conclusions. However, this pendency is secondary to the aim of the studies that are performed to optimise the prediction of (natural) conception for subfertile couples. The contribution of the male factor to the quest for the true prognosis for (natural) conception of a subfertile couple is currently scarcely investigated. It proves to be difficult to determine the prognostic capacity of the male factor for the prediction of natural conception. Not only the semen analysis (sperm motility) determines the probability of natural conception, other – female – factors have more prognostic capacity, as is illustrated by the prediction model of Hunault. In terms of fecundity the female’s contribution seems to overclass the male’s contribution. It seems unlikely with the large variability of the semen analysis that a minor improvement in the assessment of the semen analysis will exert the improvement in the prediction of natural conception one would hope for. Consecutively, it seems plausible to stimulate researchers to focus on the impact of the male factor on the prediction of natural conception. The just-mentioned MASTER study should not miss the opportunity to collect complete data on the man, such as complete history, recent illnesses and lifestyle factors as well as performance of the semen analysis in an approach that eliminates as much sources of variability as possible, i.e. preferably in a (single) laboratory that is validated with internal and external quality control schemes (EQAS) of the Dutch Foundation for Quality Assessment in Clinical Laboratories (SKML) and scores according to the WHO criteria. The unique data from this prospective study will possibly be a start of assessing the actual contribution of the male factor to the prediction of (natural) conception.
REFERENCES

21. Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. The prediction of the