INTRODUCTION
GENERAL INTRODUCTION

Prognosis is a key concept in patient care. Prognosis (Greek: πρόγνωση) literally means fore-knowing or foreseeing. A good prognostician is one who can make realistic action-specific prognoses with a high degree of objective validity and can suggest a near-optimal action. The first to incorporate the concept of prognosis as a key element in medicine was Hippocrates. Hippocrates formulated detailed criteria (prognosticators) for the prediction of death and disease.

The methodology of prognostics waned in prominence, when, after centuries of therapeutic debility effective treatments emerged, such as antibiotics in the mid-20th century. Thus, compared to diagnostic and therapeutical research, prognostic research was still relatively underdeveloped at the end of the 20th century. Now prognosis has a resurgence of interest. The availability and utility of data resources and software, and skills to build prediction models have improved considerably. It is these prediction models that allow evidence based medicine to be applied in clinical practice in a comprehensible approach. So, doctors can easily incorporate the latest evidence in their counselling and decision making.

In research, prediction models may assist the design and analysis of randomized trials. This is called Marker by Treatment Interaction Design. The design implies that the marker splits the population into groups, in which the efficacy of a particular treatment will differ. This design can be viewed as a classical randomized clinical trial with upfront stratification for the marker. Through all this, prediction models have developed into essential tools for decision-making.

In reproductive medicine, the prediction model for natural conception in subfertile couples has had a major influence on clinical care. Subfertility or involuntary childlessness occurs in approximately one out of ten couples and has a big impact on the couples involved. Similar to medicine in general, the emphasis in reproductive medicine has been on finding causal diagnoses of subfertility followed by directional treatment of the diagnosed condition. Examples are ovulation induction in women diagnosed with anovulation, tubal surgery in women with bilateral tubal disease, and in-vitro fertilization (IVF) with assisted fertilization after surgical sperm retrieval in couples with azoosperma. In many couples, such causal factors cannot be found. These couples are classified as having unexplained subfertility.

The majority of subfertile couples seek medical help. Assessment of the fertility potential in a fertility workup is then the first step. 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.

The choice between expectant management and empirical treatment, such as intrauterine insemination (IUI) or IVF, in these couples is dependent on the prognosis of the couple. Since interventions can be expensive and are not without side effects, the clinical challenge is to offer them only if the expected success rate with treatment substantially exceeds the probability of a natural conception. With a prediction model, one can calculate the probability of a treatment-independent...
pregnancy as well as the probability of success with IUI and IVF. Prediction models can also be used for starting or delaying treatment after definition of a treatment threshold or to guide decision-making once a diagnosis is made. As a consequence, validated prediction models could be a useful tool in counselling couples taking into account their individual chances of natural conception. Such prediction models could support gynecologists in making decisions on treatment such as IUI or IVF, or expectant management. From a patient’s perspective, presentation of these probabilities in a patient friendly format could encourage a couple to refrain from assisted reproduction techniques (ART) with higher costs, side effects and potentially lower chances of conception.

A single predictor or variable rarely gives an adequate estimate of prognosis. Doctors -implicitly or explicitly- use multiple predictors to estimate a patient’s prognosis in clinical practice. Prognostic studies therefore need to use a design and analysis that include multiple factors to determine the important predictors of the studied outcomes and to provide outcome probabilities for various combinations of predictors. Careful evaluation is needed before these models can be implemented in clinical practice. The use of poor-quality prediction models could have a negative effect on decision making by introducing the illusion of objective improvement over clinical judgment. Prediction models can be appraised in three phases, i.e. model derivation, model validation and impact analysis (Figure 1).

**Figure 1. Phases of model development.**

<table>
<thead>
<tr>
<th>Phase 1: Model derivation</th>
<th>Phase 2: Model validation</th>
<th>Phase 3: Impact analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of predictors and estimation of regression coefficients</td>
<td>Evidence of reproducible accuracy</td>
<td>Evidence for clinical impact by using prediction rule as a decision rule</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 2a Internal validation</th>
<th>Phase 2b External validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation of the model in the development population</td>
<td>Validation of the model in varied settings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 3a Narrow impact analysis</th>
<th>Phase 3b Broad impact analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact analysis in 1 setting</td>
<td>Impact analysis in varied settings</td>
</tr>
</tbody>
</table>

In the model derivation phase, predictors are identified based on prior knowledge and the weight of each predictor (regression coefficient) is calculated. In the second phase, the model’s predictive performance (e.g., calibration and discrimination) is validated or tested in new couples. This phase can be distinguished in internal validation (phase 2a) and external validation (phase 2b). By internal validation the model’s ability to predict outcome in the group of patients in which it was developed is evaluated, sometimes with data collected in a separate group of patients evaluated in the same setting. By external validation the model’s ability to predict outcome in populations other than the population in which the model was developed is evaluated.

As prediction models with documented validity are meant to be used for deciding on starting or delaying treatment to provide cost-effective care and assess the effects on patient outcome – the impact of prediction models after introduction into clinical
practice is evaluated in the third and final phase of impact analysis. \(^3;16;18\) This can be done in one (phase 3a) or in varied settings (phase 3b), preferably randomized controlled trials. Before this phase, facilitators and barriers for implementation have to be identified to warrant optimal adherence, i.e. to expectant management in subfertile couples with a good/intermediate prognosis of natural conception following the national Dutch Guideline for subfertility. \(^5;19;20\)

The role of the male partner in predicting conception is still unknown despite extensive research. \(^21;22\) The semen analysis is the cornerstone of the laboratory evaluation of the male partner of a subfertile couple, but the results of repeated semen analyses in a laboratory show large variation due to pre-analytic factors, such as duration of abstinence or seasonality; analytic variation in the method of analysis and/or the evaluator \(^21;23-28\); and inherent biologic variability. In addition there is a presumed large variability between laboratories. \(^22\) As a consequence, it is difficult for doctors to interpret semen analyses and to compare the results of semen analyses from different laboratories.

Since the variation of semen parameters is high, it has been proposed to deal with this by obtaining multiple semen samples rather than replicating the analysis of a given semen sample. \(^27\) In the most recent World Health Organisation (WHO) manual it is stated that it is impossible to characterize a man’s semen quality from evaluation of a single semen sample, \(^21\) but high quality evidence on how many semen analyses need to be performed during the fertility workup is lacking.

Another test for male fertility that is assumed to affect chances of pregnancy is the mixed antiglobulin reaction test in semen (direct MAR test). The WHO recommends testing for IgG by the mixed antiglobulin reaction test in semen as a routine screening method for antisperm antibodies, and if tested positive for IgG, followed by an IgA test. \(^22\) There is evidence that an abnormal test for IgG or IgA identifies couples that could benefit from intrauterine insemination (IUI). \(^29\) Although animal studies have proven that ASA affect sperm fertilizing ability at various levels, the clinical significance of IgG ASA is still not clear. Earlier studies reported conflicting results for the association of IgG ASA with natural conception. In addition, these studies all differed in design, population studied, methodology, and interpretation of results, whereas some are not compatible with current research standards.

Male and female influences on chances of pregnancy come together in the postcoital test (PCT). Since the original description of the PCT in 1866 by J. Marion Sims \(^30\) and the reintroduction by M. Huhner in 1913 \(^31\) its use has become widespread in the evaluation of subfertile couples. Many authorities consider the PCT as the cornerstone of the fertility evaluation. \(^32\) Nevertheless, the significance of this test in the basic fertility workup has been subject to debate over the last 10 years due to conflicting data. \(^33-35\) An abnormal PCT result decreases the probability of treatment-independent pregnancy twofold to threefold. \(^34;36-38\) On the other hand, routine use of the PCT would only lead to more interventions without an increase in pregnancy rates. \(^39\)
Recently, the model for the prediction of natural conception was validated in a large cohort of subfertile couples. The external validation of the model showed good predictive performance. In succession of this external validation, the addition of the PCT results to this model was evaluated. But since the study was not primarily designed to assess this and since data was derived from a substantially smaller cohort than the cohort of 2,459 couples from which the reference model was derived, valid conclusions for evaluation of the prognostic value of the PCT could not be made based on this analysis.

BACKGROUND OF THE RESEARCH DESCRIBED IN THIS THESIS

It is the introduction of the concept of prognosis in reproductive medicine, the availability of new knowledge and utility of new data, that opens a window of opportunities to evaluate prediction models. In this thesis we focus on male subfertility and prediction of natural conception.

First, we review the literature on the available prediction models in reproductive medicine. The focus of the review is to appraise the prediction models according to the afore mentioned evaluation scheme of phases of model development. Second, we evaluate the value of the semen analysis and postcoital test from the perspective of predicting natural conception.

Several fertility guidelines recommend to repeat the semen analysis once or twice in an attempt to achieve a more reliable approximation of the true values of individual semen parameters. But high quality evidence on how many semen analyses need to be performed during the fertility workup to predict natural conception is lacking. Yet, the first step in deciding whether or not to repeat the semen analysis in the basic fertility workup is to assess the actual degree of within-subject variability that is represented by the reproducibility and reliability in male partners of subfertile couples. Therefore we perform a retrospective cohort study to establish the precise degree of variability in male partners of subfertile couples.

We then perform a prospective cohort study to evaluate whether or not two semen analyses predict natural conception better than one semen analysis. Results of the semen analysis can vary considerably between laboratories. Standardization of semen analysis results could improve the reproducibility of the test. We evaluate if systematic differences between laboratories exist and if Z-score and regression transformations can be used to standardize semen analysis results to reduce such differences or to obligate the repetition of the semen analysis when referral to another hospital takes place.

We 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. Finally, we assess the prognostic value of the PCT in addition to the established model for the prediction of natural conception. We design and perform a study to evaluate the prognostic value of the PCT in a large prospective multicenter cohort of subfertile couples in relation to the existing well validated prediction model for natural conception.
OUTLINE OF THE THESIS

In Chapter 2 we provide a review on the current literature on the available prediction models in reproductive medicine for three strategies: expectant management, intrauterine insemination (IUI) or in vitro fertilization (IVF). We appraise the prediction models in three phases, i.e. model derivation, model validation and impact analysis and summarize their performance at external validation in terms of discrimination and calibration.

In Chapter 3 we report the actual degree of within-subject variability in male partners of subfertile couples, represented by reproducibility and reliability. We perform a retrospective cohort study in two university hospitals in The Netherlands, which routinely perform two semen analyses in the male partner of subfertile couples. We assess the test-retest reproducibility, by calculating the coefficient of variation ($CV_w$) and the reliability (in terms of the intraclass correlation coefficient (ICC)) for five semen parameters in a cohort of 5,240 men visiting the two hospitals between January 1998 and 2008.

In Chapter 4 we first evaluate whether two semen analyses predict natural conception better than one semen analysis in a prospective cohort study of 897 men of consecutive couples presenting for subfertility in the period 2002 to 2004. Second, we 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. We calculate associations between the results of the two routinely collected semen analysis and natural conception within a time horizon of one year. Based on three semen parameters, three strategies with models for the prediction of natural conception will be constructed, by Cox univariable and multivariable regression analyses, respectively based on a single semen analysis, two semen analyses and taking the average as the final result, and a second semen analysis only if the TMC of the first semen analysis is below $10^6$, again taking the average in that case.

In Chapter 5 we assess systematic differences as well as the variability between laboratories and evaluate whether a transformation using Z-scores and regression statistics can be used to standardize semen analysis results and to reduce systematic differences between laboratories in semen analysis results. We scored semen parameters from semen samples of 8 men that circulated between 12 laboratories to calculate the transformation to the Z-score and regression coefficients for each laboratory from these scores. We use the Z-score and the regression coefficient transformations to standardize the semen analysis results from a large cohort study of 2,804 men and repeat the test for assessment of systematic differences between laboratories. Concurrently, we calculate the between-laboratory coefficient of variation ($CV_b$).

In Chapter 6 we investigate the capacity of immunoglobulin G (IgG) antisperm
antibodies (ASA) of the semen analysis to predict natural ongoing pregnancy in ovulatory subfertile couples. We perform a large prospective cohort study of 1,794 couples from nine fertility centers in The Netherlands between January 2002 and February 2004.

In Chapter 7 we report on the capacity of the postcoital test (PCT) to predict natural conception in a large prospective cohort study of 3,021 subfertile couples from the department of reproductive medicine of 38 hospitals in the Netherlands between January 2002 and February 2004. We estimate the contribution of the PCT result to the existing prediction model for natural conception by calculating the adjusted hazard ratio (HR) of an abnormal PCT result. We construct a second prediction model (PCT model) based on the reference model including the PCT and evaluate the performance of the PCT model in comparison with the reference model by calculating goodness of fit, discrimination, calibration, and the “net reclassification improvement”.

In Chapter 8 we give a summary of this thesis and provide implications for future research.
REFERENCES

21. World Health Organization. WHO laboratory manual for the examination and processing of
CHAPTER 1 INTRODUCTION


