Predicting IVF outcome

van Loendersloot, L.L.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Predicting IVF Outcome

On 25 July 1978 at 11.47 PM Louise Brown was born as the first IVF baby ever. Since its introduction more than 5 million babies have been born worldwide using IVF. In contrast to patients’ perception, IVF does not guarantee success; almost 50% of couples that start with IVF will not achieve a pregnancy through IVF even if they undergo multiple cycles. Given this limited success, it seems logical to offer IVF only to couples with reasonable chances of success and to discontinue treatment when chances are low and do not outweigh the burden and costs associated with treatment. As doctors are not able to correctly predict these pregnancy chances, prediction models can be a useful tool.

Another concern in current IVF practice is the high multiple pregnancy rates as multiple pregnancies are associated with an increase in maternal and perinatal morbidity and mortality as well as costs. A more individualized embryo transfer strategy could be a solution.

This PhD thesis describes the development and validation of several prediction models in IVF. The first part of this thesis focuses on couples’ prognosis with IVF. The second part of this thesis focuses on optimizing embryo transfer strategies.
Predicting IVF Outcome

Laura van Loendersloot
Predicting IVF Outcome

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college
voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 3 december 2013, te 16:00 uur

door

Laura Lotte van Loendersloot

geboren te Laren
Promotiecommissie

Promotores:  
Prof. dr. F. van der Veen  
Prof. dr. P.M.M. Bossuyt

Co-promotores:  
Dr. M. van Wely  
Prof. dr. S. Repping

Overige leden:  
Prof. dr. S. Bhattacharya  
Dr. J.C.M. Dumoulin  
Prof. dr. C.B. Lambalk  
Prof. dr. B.W.J. Mol  
Prof. dr. E.W. Steyerberg

Faculteit der Geneeskunde
Aan mijn ouders,
Marion & Nick
## CONTENTS

Chapter 1  Introduction  9

**Part one:  Prognosis with in vitro fertilization**


Chapter 5  Prediction models in in vitro fertilization; where are we? *Journal of Advanced Research* 2013; *in press: May 9*  83

**Part two:  Optimizing embryo transfer strategies**


Chapter 7  Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential using morphological scoring *Reproductive BioMedicine Online conditionally accepted September 2013*  117

Chapter 8  Balancing pregnancy and multiple pregnancy chances: an embryo transfer model. *Manuscript in preparation*  139

Chapter 9  Summary and implications for further research  159

Chapter 10  Samenvatting en implicaties voor toekomstig onderzoek  169

Appendices  Portfolio  181

Dankwoord  187

About the author  193
Introduction
Chapter 1

On 25 July 1978 at 11.47 PM Louise Brown was born as the first IVF baby ever. Since then, the number of in vitro fertilization cycles has increased rapidly: in the United Kingdom 6,650 cycles were performed in 1991 and 61,726 cycles in 2011 (1,2). In the Netherlands, 11,154 cycles were performed in 1996 and 16,668 cycles in 2011 (3). Since its introduction more than 5 million babies have been born worldwide using IVF (4). IVF is currently one of the most widely used interventions for infertility.

The increase in the number of IVF cycles is not caused by a sudden epidemic of infertility, but mainly by increased access to IVF and by expansion of the indications for IVF. At first, IVF was initiated in couples with bilateral tubal occlusion (5). In 1992 intracytoplasmic sperm injection (ICSI) was initiated in couples with severe male subfertility. Later on IVF was also applied in couples with unexplained subfertility, cervical hostility, failed ovulation induction, endometriosis, or unilateral tubal pathology (6-8). The major difference between the original indication and the indications for which IVF is conducted nowadays, is that the couples with bilateral tubal pathology or azoospermia have zero chance of natural conception and completely depend on IVF/ICSI for getting pregnant, whilst couples with other indications are subfertile, and do have chances of natural conception, which may or may not be better than with IVF. For them, these chances have to be balanced against those with IVF.

In contrast to patients’ perception, IVF does not guarantee success; almost 50% of couples that start with IVF will not achieve a pregnancy through IVF even if they undergo multiple cycles (9). Given this limited success, it seems logical to offer IVF only to couples with reasonable chances of success and to discontinue treatment when chances are low and do not outweigh the burden and costs associated with treatment. Unfortunately, evidence from randomised trials confirming the effectiveness of IVF over natural conception in couples without bilateral tubal pathology or severe male factor is scarce. In fact only one single, small trial of only 51 couples has compared in couples with unexplained subfertility one cycle of IVF to three months of natural conception (10). The live birth rate was significantly higher with a single cycle of IVF (RR 12.4, 95% CI 1.72-89) (11). Due to paucity of data from randomised trials the effectiveness of IVF relative to natural conception remains unproven. So, until larger randomized trials with a prolonged duration of follow-up (more than one cycle), and with sufficient power are performed, the only way to prevent overtreatment and to counsel couples properly is selection based on the couple’s prognosis.

Unfortunately, gynecologists are not able to estimate the probability of achieving a pregnancy with IVF accurately (12). To support counseling, patient selection and clinical decision making in IVF, a number of prediction models have been developed in the past (7,13). Several models are of limited use since they were developed before current clinical and laboratory protocols were established (14-24). Most models do not include the transfer of frozen-thawed embryos, an essential component of modern day IVF.
Introduction

(14-32). A number of models calculate pregnancy chances only for the first IVF cycle, whilst other models calculate pregnancy chances only after one or more failed IVF cycle (14;16;20;26;32-35).

The decisions to start or continue IVF are now largely guided by expert opinion. This approach ignores predictive factors other than female age and is not refined enough for clinical practice. It may lead to futile treatment or to wrongfully refraining from treatment.

The second major concern in current IVF practice is the high multiple pregnancy rate. In reproductive medicine the goal of every fertility treatment is live birth. Until a decade ago, one of the most common approaches to increase the likelihood of pregnancy in IVF was to transfer multiple embryos into the uterine cavity (36). This approach indeed resulted in high pregnancy rates but also increased the risk of multiple pregnancies. Initially, a multiple pregnancy was felt to be justified in what was largely an experimental treatment with poor pregnancy and live birth rates (37). In the last two decades pregnancy rates have increased substantially but multiple pregnancy rates remained high (38-40). The high multiple pregnancy rates caused concern, since multiple pregnancies are associated with an increase in maternal and perinatal morbidity and mortality as well as costs (41-44).

To curtail the multiple pregnancy rates, several treatment strategies have been evaluated over the years. The first strategy was to reduce triple embryo transfers and to increase double embryo transfers (45;46). This indeed prevented most triplets, but did not diminish the rate of twin pregnancies (39). As a result, single embryo transfer was introduced. The first randomized trial to compare elective single embryo transfer (eSET) with double embryo transfer (DET) was performed in 1999 in Belgium in women under 34 undergoing their first IVF/ICSI cycle (37). Since then, two systematic reviews and meta-analyses of randomized trials have been performed, based on synthesized aggregated data and individual patient data on eSET versus double embryo transfer (DET) in women with a good prognosis, i.e. women younger than 36 years and with at least two good quality embryos. These meta-analyses showed that, although eSET minimized the odds of multiple pregnancies, it also halved the odds of a live birth per fresh cycle. Subsequent transfer of a single frozen thawed embryo resulted in comparable cumulative live birth rates to those after DET (47;48). Although eSET is now an accepted policy for women with a good prognosis, the majority of women currently undergoing IVF have an intermediate or poor prognosis, such as women over 35 with several failed IVF cycles. For these women it is less clear how many embryos have to be transferred to obtain high pregnancy rates at low multiple pregnancy rates.
Chapter 1

BACKGROUND OF THE RESEARCH OF THIS THESIS

At the start of the studies described in this thesis, there was only one IVF prediction model with good predictive performance that could be used in clinical practice (24). Yet, an IVF prediction model that predicts pregnancy chances during the complete IVF procedure, after failed cycles, and after fresh and frozen-thawed embryo transfer, did not exist. Also, if one uses a model and embarks on IVF, the ultimate goal is a singleton, since multiple pregnancies are associated with higher maternal and neonatal morbidity and mortality. At the start of the work reported in this thesis multiple pregnancy rates associated with IVF were still too high, and it was not clear how many embryos should be transferred in the general IVF population to obtain high pregnancy rates at low multiple pregnancy rates. Optimization of embryo transfer strategies was thus necessary.

We first externally validated the Templeton prediction model developed in 1996 in a more recent dataset, as IVF has progressed substantially during the last two decades we thought it was questionable whether the model was still valid in current clinical practice.

We then performed a systematic review and meta-analysis of the literature to identify candidate predictive factors and developed, based upon the identified factors, an IVF prediction model that could calculate pregnancy chances prior to the start of the first IVF cycle as well as after one or more failed cycles, and one that would take into account pregnancy chances after both fresh- and frozen-thawed embryo transfers.

We also compared different strategies to optimize embryo transfer and to reduce the multiple pregnancy rates whilst maintaining optimal pregnancy rates. In a prospective cohort we evaluated the results of a differentiated embryo transfer policy based on the prognostic profile of a woman, i.e. female age, cycle number and embryo quality.

We then developed an individualized embryo transfer strategy by building an embryo implantation model that ranks embryos based on their implantation potential and extended this model to an embryo transfer model that is not only able to calculate pregnancy chances after IVF, but also the chances of a single- or multiple pregnancy after single-, double, or triple embryo transfer.

OUTLINE OF THIS THESIS

Part one: Prognosis with in vitro fertilization

Chapter 2 presents an external validation of the IVF model developed by Templeton et al. in 1996. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic (ROC) curve (AUC), and calibration. We evaluated likely causes for miscalibration by refitting the Templeton model to our data.
Introduction

Chapter 3 provides a systematic review and meta-analysis of published literature on nine predictive factors for success in IVF: age, type of infertility, indication, duration of infertility, basal FSH, number of oocytes, fertilization method, number of embryos transferred and embryo quality. Fourteen studies were identified. For five factors a summary odds ratio is presented.

Chapter 4 reports on the development of a new IVF prediction model that is able to calculate pregnancy chances prior to the start of the first IVF cycle as well as after one or more failed cycles. We collected data on couples who had been treated with IVF or ICSI between January 2001 and July 2009 at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam. A model was developed using multivariable logistic regression and a blockwise model building strategy to calculate the probability of a pregnancy with IVF, including fresh and frozen-thawed embryo transfers from the same cycle. The model was validated in additional data from couples treated between August 2009 and April 2011 at the same center.

Chapter 5 provides an overview on predictive factors in IVF, the available prediction models in IVF and provides key principles that can be used to critically appraise the literature on prediction models in IVF.

Part two: Optimizing embryo transfer strategies
In the prospective cohort study reported in Chapter 6 we evaluated in a tailored embryo transfer policy based on the prognostic profile of the couple. Between August 2006 and April 2011 we adhered to the following embryo transfer protocol at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam: single embryo transfer (SET) was performed followed by double embryo transfer (DET) in frozen embryo transfer cycles in women with a good prognosis (aged <35 years, first cycle, ≥1 top quality embryo). DET was performed in both fresh and frozen cycles in women with an intermediate prognosis (<35 years, first cycle and no top quality embryo available, or aged <35 years and ≥1 failed cycles, or aged 35-38 years). Triple embryo transfer (TET) in both fresh and frozen cycles was performed in women with a poor prognosis (aged ≥ 39 years).

Chapter 7 reports on the development of a prediction model to rank embryos within a single IVF/ICSI cycle according to their ongoing implantation potential. We prospectively studied embryo transfers on day 3 between January 2004 and July 2009 at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam. We evaluated pronuclear score, early cleavage, number of blastomeres on day 2 and day 3, morphological score on day 2 and day 3 as potential predictors for implantation. A model was developed using multivariable logistic regression. The prediction model was externally, temporally validated on embryo transfer data between August 2009 and September 2011 from the same center.
Chapter 1

Chapter 8 reports on the development of an embryo transfer model that is able to calculate the probability of an ongoing pregnancy as well as the probability of multiple pregnancy after single-, double embryo or triple embryo transfer. We collected data on couples who had been treated with IVF or ICSI between January 2004 and April 2011 at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam. The embryo transfer model consists of two components: variables specific to the transferred embryo(s), such as the number of blastomeres and variables specific to the couple, including maternal and treatment factors that affect all embryos equally. The model was based on the two previously developed models described in chapters 4 and 7. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic curve (AUC) and calibration.

Chapter 9 summarizes the data presented in this thesis with implications for further research.
Introduction

REFERENCES


Chapter 1


Introduction


PART ONE

Prognosis with in vitro fertilization
The Templeton prediction model underestimates IVF success in an external validation

L.L. van Loendersloot
M. van Wely
S. Repping
F. van der Veen
P.M.M. Bossuyt

Reproductive BioMedicine Online 2011; 22:597-602
ABSTRACT

Prediction models for in vitro fertilization (IVF) can be used to identify couples that will benefit from IVF treatment. Currently there is only one prediction model with a good predictive performance that can be used for predicting pregnancy chances after IVF. That model was developed almost 15 years ago and since IVF has progressed substantially during the last two decades it is questionable whether the model is still valid in current clinical practice. The objective of our study was to validate the prediction model of Templeton for calculating pregnancy chances after IVF. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic (ROC) curve (AUC) and calibration. We evaluated likely causes for miscalibration by refitting the Templeton model to our data. The area under the ROC curve for the Templeton model was 0.61. Calibration showed a significant and systematic underestimation of success in IVF. Although the Templeton model can distinguish somewhat between women with a high and those with a low success rate in IVF, it systematically underestimates pregnancy chances and has therefore no real value for current IVF practice.
INTRODUCTION

Louise Brown, the first vitro fertilisation (IVF) baby, was born in 1978. Initially, IVF was used to bypass infertility in women with bilateral tubal occlusion (1). In later years, the indication for IVF broadened to couples with unexplained subfertility, male subfertility, cervical factor, failed ovulation induction, endometriosis, or unilateral tubal pathology (2-4). Currently IVF is a widely used treatment for infertility.

IVF is not effective in all couples, in fact only 50% of couples will have a live born baby (5). Because IVF is expensive and has side effects it should not be offered to couples with conception chances that approach zero. To help gynaecologists in their clinical decision making and to facilitate patient counselling, prediction models have been developed (6). A recent study showed that there is currently only one prediction model with a good predictive performance that can be used for predicting pregnancy chances after IVF (6;7). The authors suggest that this model could be used as a guiding tool in making decisions about IVF treatment in subfertile couples. Since IVF techniques have evolved since the Templeton model was developed and IVF pregnancy rates have increased during the past decade, it is questionable whether the Templeton model is still valid today. The aim of this study was to evaluate the performance of the Templeton prediction model using recently collected IVF data.

MATERIALS AND METHODS

We evaluated the Templeton model using data on patient characteristics and IVF outcome prospectively collected in the Centre for Reproductive Medicine of the Academic Medical Centre, the Netherlands between January 2001 and September 2009.

For the first external validation we mirrored the design of the Templeton study, the following cycles were excluded from our analysis: cycles that involved oocyte or embryo donation, sperm obtained by testicular extraction, cycles with intracytoplasmatic spermation injection (ICSI), cycles from HIV positive patients, cycles with frozen embryo transfer and cycles that were unstimulated (natural IVF). All cycles were included once ovarian hyperstimulation had started. The stimulation protocol included ovarian down-regulation with a gonadotrophin-releasing hormone agonist starting in the midluteal phase. Ovarian hyperstimulation was conducted with recombinant follicle-stimulating hormone or human menopausal gonadotrophin.

We performed a second validation in the set of all cycles with ICSI and excluding all regular IVF cycles.
Chapter 2

Analysis
We evaluated the model in its ability to predict an ongoing pregnancy with IVF, which was defined as the presence of foetal cardiac activity seen at transvaginal ultrasound at a gestational age of at least 10 weeks. The predicted probability ($P$) of achieving a pregnancy after IVF was calculated using the Templeton model:

$$P = \frac{1}{1+e^{-y}}$$

Where $y$ was defined as $y = -2.028 + 0.00551 \times (Age - 16)^2 - 0.00028 \times (Age - 16)^3 + (i) - 0.0690 \times \text{no. of unsuccessful IVF attempts} - 0.0711 \times \text{tubal infertility} + 0.7587 \times \text{ongoing pregnancy after IVF} + 0.2986 \times \text{previous pregnancy after IVF which did not result in ongoing pregnancy} + 0.2277 \times \text{ongoing pregnancy which was not a result of IVF} + 0.1117 \times \text{previous pregnancy, not after IVF and which did not result in ongoing pregnancy}.$

The indicator $i$ was a value used to represent the duration of infertility in years and was 0.2163 if the duration of infertility was 1–3 years, −0.0839 if duration was 4–6 years, −0.1036 if duration was 7–12 years, and −0.4179 if duration of infertility was ≥13 years.

Model performance was expressed in terms of its discrimination and calibration. Discriminative capacity was assessed by estimating the area under the receiver operation characteristic (ROC) curve for the model.

Discrimination, as expressed by c-statistics or area under ROC curve, does not express very well the extent to which predictive models can guide decision making (8). Calibration is a more informative way of summarizing the performance of a model (8). Calibration expresses whether the calculated probabilities agree with the observed relative frequencies (9-11). Calibration was assessed by comparing the average calculated probabilities of an ongoing pregnancy rate with the observed ongoing pregnancy rate in disjoint subgroups. For this purpose, our cohort was split into 5 groups based on the quintiles of the calculated cycle-based probabilities of achieving a pregnancy. In each group, the average calculated probability and the observed pregnancy rate were calculated and compared with one another. We positioned the average probabilities and corresponding observed pregnancy rates, with their 95% confidence intervals, as points in a calibration plot. In case of perfect calibration, all points in a calibration plot are on the diagonal, the line of equality, and probabilities correspond perfectly to the actual rates.

We also calculated the intercept and the slope for the regression line in the calibration plot. The intercept can be interpreted as the extent to which predictions are systematically too low or too high, i.e. ‘calibration-in-the-large’. The calibration slope, if smaller than 1, shows that the model generates overoptimistic predictions: low probabilities are too low and high probabilities are too high. In case of a perfect calibration, the intercept will be zero and the slope would be one.
The Templeton prediction model underestimates IVF success in an external validation

In addition, we refitted the Templeton model on our data and compared the estimates of the coefficients for the factors in the model with the respective estimates reported from the Templeton development study. This was done to evaluate likely causes for any miscalibration and to assess the predictive value of the variables, excluding any effects from differences between the development population and our study group. After evaluating the model on the IVF data, we also evaluated it on ICSI data. All calculations were performed using the Statistical Package for Social Sciences (SPSS) version 15.0 (SPSS Inc., Chicago).

RESULTS

We included data from 1394 couples, who underwent 1,537 cycles of IVF and 1442 cycles of ICSI. Baseline characteristics and pregnancy outcome of these couples are summarized in Table I. The average female age of women undergoing IVF was 35.9 years and of women undergoing ICSI 34.4 years. In total, 383 pregnancies occurred, of which 262 were ongoing (17% per cycle).

Table I | Baseline characteristics and pregnancy outcome after IVF for the AMC cohort and Templeton cohort

<table>
<thead>
<tr>
<th></th>
<th>AMC</th>
<th>ICSI</th>
<th>Templeton</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IVF % of cycles</td>
<td>ICSI % of cycles</td>
<td>% of cycles</td>
</tr>
<tr>
<td>Number of patients</td>
<td>688</td>
<td>706</td>
<td>26389</td>
</tr>
<tr>
<td>Number of IVF cycles</td>
<td>1537</td>
<td>1442</td>
<td>36961</td>
</tr>
<tr>
<td>Number of ongoing pregnancies</td>
<td>262</td>
<td>17.0%</td>
<td>316</td>
</tr>
<tr>
<td>Female age (SD)</td>
<td>35.9 (4.3)</td>
<td>34.4 (4.8)</td>
<td></td>
</tr>
<tr>
<td>Duration of subfertility (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>66</td>
<td>4.3%</td>
<td>79</td>
</tr>
<tr>
<td>1-3</td>
<td>876</td>
<td>57.0%</td>
<td>891</td>
</tr>
<tr>
<td>4-6</td>
<td>437</td>
<td>28.4%</td>
<td>355</td>
</tr>
<tr>
<td>7-9</td>
<td>110</td>
<td>7.2%</td>
<td>76</td>
</tr>
<tr>
<td>10-12</td>
<td>24</td>
<td>1.6%</td>
<td>25</td>
</tr>
<tr>
<td>&gt;12</td>
<td>24</td>
<td>1.6%</td>
<td>16</td>
</tr>
<tr>
<td>Case of infertility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal disease</td>
<td>558</td>
<td>36.3%</td>
<td>88</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>104</td>
<td>6.8%</td>
<td>25</td>
</tr>
<tr>
<td>Unexplained</td>
<td>463</td>
<td>30.1%</td>
<td>110</td>
</tr>
<tr>
<td>Cervical and uterine</td>
<td>65</td>
<td>4.2%</td>
<td>22</td>
</tr>
</tbody>
</table>
Chapter 2

Continuation of table I

<table>
<thead>
<tr>
<th></th>
<th>AMC IVF</th>
<th></th>
<th>AMC ICSI</th>
<th></th>
<th>Templeton</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of cycles</td>
<td></td>
<td>% of cycles</td>
<td></td>
<td>% of cycles</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>291</td>
<td>18.9%</td>
<td>1187</td>
<td>82.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>292</td>
<td>19.0%</td>
<td>205</td>
<td>14.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Numer of previous treatment cycles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>898</td>
<td>58.4%</td>
<td>581</td>
<td>40.3%</td>
<td>18239</td>
<td>49.3%</td>
</tr>
<tr>
<td>1</td>
<td>396</td>
<td>25.8%</td>
<td>438</td>
<td>30.4%</td>
<td>8123</td>
<td>22.0%</td>
</tr>
<tr>
<td>2</td>
<td>167</td>
<td>10.9%</td>
<td>252</td>
<td>17.5%</td>
<td>3706</td>
<td>10.0%</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>3.3%</td>
<td>118</td>
<td>8.2%</td>
<td>1786</td>
<td>4.8%</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>1.2%</td>
<td>36</td>
<td>2.5%</td>
<td>864</td>
<td>2.3%</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.4%</td>
<td>11</td>
<td>0.8%</td>
<td>389</td>
<td>1.1%</td>
</tr>
<tr>
<td>6-9</td>
<td>5</td>
<td>0.3%</td>
<td></td>
<td></td>
<td>483</td>
<td>1.3%</td>
</tr>
<tr>
<td>≥10</td>
<td>1</td>
<td>0.1%</td>
<td></td>
<td></td>
<td>111</td>
<td>0.3%</td>
</tr>
<tr>
<td><strong>Number of previous pregnancies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>956</td>
<td>62.2%</td>
<td>943</td>
<td>65.4%</td>
<td>19997</td>
<td>54.1%</td>
</tr>
<tr>
<td>≥1 not by IVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No previous livebirth</td>
<td>251</td>
<td>16.3%</td>
<td>101</td>
<td>7.0%</td>
<td>8388</td>
<td>22.7%</td>
</tr>
<tr>
<td>≥1 previous livebirth</td>
<td>197</td>
<td>12.8%</td>
<td>105</td>
<td>7.3%</td>
<td>5336</td>
<td>14.4%</td>
</tr>
<tr>
<td>≥1 by IVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No IVF livebirth</td>
<td>102</td>
<td>6.6%</td>
<td>223</td>
<td>15.5%</td>
<td>1786</td>
<td>4.8%</td>
</tr>
<tr>
<td>≥1 IVF livebirth</td>
<td>153</td>
<td>10.0%</td>
<td>184</td>
<td>12.8%</td>
<td>1451</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

Table II | Calibration: calculated probability of an ongoing pregnancy versus observed pregnancy rates after IVF

<table>
<thead>
<tr>
<th>Probability Range</th>
<th>No. of patients in group</th>
<th>Average Probability</th>
<th>No. of pregnancies</th>
<th>Pregnancy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% - 7%</td>
<td>311</td>
<td>5.2%</td>
<td>34</td>
<td>10.9%</td>
</tr>
<tr>
<td>7% - 10%</td>
<td>301</td>
<td>8.8%</td>
<td>38</td>
<td>12.7%</td>
</tr>
<tr>
<td>10% - 13%</td>
<td>311</td>
<td>11.8%</td>
<td>54</td>
<td>17.4%</td>
</tr>
<tr>
<td>13% - 17%</td>
<td>306</td>
<td>14.8%</td>
<td>46</td>
<td>15.1%</td>
</tr>
<tr>
<td>&gt;17%</td>
<td>308</td>
<td>20.0%</td>
<td>90</td>
<td>29.2%</td>
</tr>
</tbody>
</table>

The discrimination obtained with the Templeton model between couples with and couples without an ongoing pregnancy that underwent IVF only resulted in an area under the ROC curve of 0.61 (95% confidence interval: 0.57 to 0.65) (Figure 1). The average probabilities of an ongoing pregnancy after IVF, calculated with the Templeton model, compared with the observed ongoing pregnancy rates, are summarized in Table II and Figure 2.
The Templeton prediction model underestimates IVF success in an external validation

Figure 1 | Receiver Operating Characteristic - curve for IVF only

The difference between the calculated probabilities and observed pregnancy rates varied between 0.3% and 9.2% in the five subgroups, with the observed ongoing pregnancy rates always being higher than the corresponding probabilities (Table II). The model seemed to be well calibrated for ongoing pregnancy probabilities in the range between 13% and 17%; in all other subgroups there was a systematic underestimation (Figure 2). The confidence intervals of the group with good probabilities (>17%) did not overlap with the confidence intervals of the other groups indicating a significant distinction between these prognostic groups (Figure 2).

The calibration-in-the-large coefficient for the Templeton model – the intercept of the regression line in the calibration plot - was 0.08, also indicating that predictions are systematically too low. The estimated slope of the regression curve was 0.74, demonstrating an overoptimism of around 26%, which means that low pregnancy rates calculated by the model are too low and high pregnancy rates calculated by the model are too high compared to the observed pregnancy rates.

After refitting the model, the coefficients for age, the first 3 categories of duration of infertility, number of previous unsuccessful IVF attempts, tubal pathology, previous (ongoing) pregnancies after IVF and ongoing pregnancy (excluding IVF pregnancy) were similar to the ones in the original Templeton model (Table III). The coefficients for the variables previous pregnancy not resulting in ongoing pregnancy (excluding IVF pregnancy) and 13 years or more years of infertility had an opposite sign.
Chapter 2

**Figure 2** | Calibration plot, showing the association between the calculated and observed rates of ongoing pregnancy after IVF.

![Calibration plot, showing the association between the calculated and observed rates of ongoing pregnancy after IVF](image)

**Table III** | Refitting the Templeton model on the AMC data set for IVF.

<table>
<thead>
<tr>
<th></th>
<th>Templeton</th>
<th>AMC</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>OR</td>
<td>Lower</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age*</td>
<td>1.006</td>
<td>1.004</td>
<td>0.996</td>
</tr>
<tr>
<td>Age†</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Duration of infertility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>1.24</td>
<td>1.01</td>
<td>0.50</td>
</tr>
<tr>
<td>4 years</td>
<td>0.92</td>
<td>0.68</td>
<td>0.32</td>
</tr>
<tr>
<td>7 years</td>
<td>0.90</td>
<td>0.60</td>
<td>0.25</td>
</tr>
<tr>
<td>13 years</td>
<td>0.66</td>
<td>1.12</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Number of previous unsuccessful IVF attempts</strong></td>
<td>0.93</td>
<td>0.82</td>
<td>0.69</td>
</tr>
<tr>
<td><strong>Tubal reason for infertility</strong></td>
<td>0.93</td>
<td>0.98</td>
<td>0.73</td>
</tr>
<tr>
<td>IVF ongoing pregnancy</td>
<td>2.14</td>
<td>1.15</td>
<td>0.70</td>
</tr>
<tr>
<td>IVF pregnancy not ongoing pregnancy</td>
<td>1.35</td>
<td>2.52</td>
<td>1.47</td>
</tr>
<tr>
<td>Ongoing pregnancy (ex. IVF)</td>
<td>1.26</td>
<td>1.38</td>
<td>0.93</td>
</tr>
<tr>
<td>Not ongoing pregnancy (ex. IVF)</td>
<td>1.12</td>
<td>0.87</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>0.13</td>
<td>0.29</td>
<td></td>
</tr>
</tbody>
</table>

Age* is calculated as (Age-16)^2
Age† is calculated as (Age-16)^3
The Templeton prediction model underestimates IVF success in an external validation.

**Figure 3** | Calibration plot, showing the association between the calculated and observed rates of ongoing pregnancy after ICSI.

| Table IV | Refitting the Templeton model on the AMC data set for ICSI. |
| --- | --- | --- | --- |
| | Templeton OR | AMC OR | 95% C.I. |
| Age | | | Lower | Upper |
| Age* | 1.006 | 1.006 | 1.000 | 1.012 |
| Age† | 1.000 | 1.000 | 0.999 | 1.000 |
| Duration of infertility | | | | |
| 1 year | 1.24 | 0.98 | 0.56 | 1.71 |
| 4 years | 0.92 | 1.00 | 0.53 | 1.88 |
| 7 years | 0.90 | 0.81 | 0.38 | 1.75 |
| 13 years | 0.66 | 1.14 | 0.21 | 6.13 |
| Number of previous unsuccessful IVF attempts | 0.93 | 0.88 | 0.77 | 1.01 |
| Tubal reason for infertility | 0.93 | 1.08 | 0.63 | 1.86 |
| Previous pregnancy | | | | |
| IVF ongoing pregnancy | 2.14 | 1.80 | 1.20 | 2.68 |
| IVF pregnancy not ongoing pregnancy | 1.35 | 0.87 | 0.57 | 1.31 |
| Ongoing pregnancy (ex. IVF) | 1.26 | 1.44 | 0.88 | 2.35 |
| Not ongoing pregnancy (ex. IVF) | 1.12 | 0.94 | 0.54 | 1.61 |
| Intercept | 0.13 | 0.27 | |

Age* is calculated as (Age-16)^2
Age† is calculated as (Age-16)^3
When calibrating the model on ICSI cycle data, the model also systematically underestimated pregnancy chances (Figure 3). After refitting the model on the ICSI datasets, the coefficients for the variables previous pregnancy not resulting in ongoing pregnancy (excluding IVF pregnancy) and 13 years or more years of infertility had an opposite sign as well compared to the Templeton dataset. In addition duration of infertility of one year had an opposite sign as well (Table IV).

**DISCUSSION**

We assessed the validity of the Templeton model, which calculates ongoing pregnancy chances after IVF. Our study showed that the Templeton model had moderate discrimination but poor calibration in more recently collected data. The Templeton model was able to differentiate to some extent between women with low pregnancy chances from women with high pregnancy chances. Yet using this model also led to a systematic underestimation of the pregnancy chances after IVF and also after ICSI. All probabilities calculated with the Templeton model were substantially and significantly lower than the observed pregnancy rates after IVF.

The performance of the model has previously been evaluated in another study (12). This validation study also showed that the model had a poor discriminative capacity with a c-static of 0.63 (12). However in contrast to our results, the calibration of the model was good (6;12). The poor calibration of the model in our study with a more recent dataset, is most likely caused by the changes in clinical practice in ART. Pregnancy rates for IVF have steadily increased during the last decade and new techniques such as intracytoplasmic sperm injection (ICSI) were introduced (13;14). The relatively small number of cycles in our study could also be responsible for a poorer calibration of the model.

Another possible explanation for the poor calibration could be a change in the patient population referred to fertility clinics. The number of IVF cycles has increased enormously and nowadays IVF has become a widely accepted treatment (13;14). Because of this, IVF has also become more accessible to women. It is possible that nowadays patients are referred sooner to a fertility clinic than before, resulting in a shorter average duration of infertility. On the other hand, there is an increasing number of women postponing childbearing (15-17). As age is one of the most important predictive factors for success in IVF and IVF cannot compensate for age related infertility, the increasing age of women results in a subfertile population with a poorer prognosis at the same duration of infertility than a decade ago (18;19).

Also the patient profile in our study differs from the Templeton study. In our centre we used a classification of six indications for IVF, instead of four. This is obviously a consequence of the broadening of the indication for IVF over the years and the
The Templeton prediction model underestimates IVF success in an external validation

introduction of assisted fertilization with ejaculated or surgically retrieved sperm in couples with severe male subfertility. The IVF population of 2010 is therefore likely to differ from the IVF population in the Templeton database. All of this may have resulted in a reduced applicability and poor calibration.

In this analysis, we used ongoing pregnancy as an outcome measure instead of live birth. Approximately 1-2% of all ongoing pregnancies can result in a late miscarriage or still birth (20). Using a different outcome measure could potentially cause bias, but since only a very small number of ongoing pregnancies do not result in a live birth, we think that this would not fundamentally alter our results.

Although the model does not perform well in a recent dataset, the predictive factors included in the model are shown to still be predictive today. Adding other new predictive factors such as ovarian reserve test, type of fertilisation, results of ovarian stimulation, number of oocytes, number of embryos, embryo quality and also cryopreserved/thawed embryos to the model may increase its performance.

In summary, our results show that the combination of factors, identified by Templeton and colleagues fifteen years ago, is still predictive today. These predictive factors allow us to distinguish between women with low and women with high chances of a pregnancy with IVF. Nevertheless, the calculated probabilities are too low, and this can lead to errors in decision making, like refraining from IVF treatment based on underestimated chances of success. Therefore the model has no real value for current IVF practice. A new model, developed on a more recent data set and possibly including additional predictive factors, is highly warranted.
REFERENCES

The Templeton prediction model underestimates IVF success in an external validation.

Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

L.L. van Loendersloot
M. van Wely
J. Limpens
P.M.M. Bossuyt
S. Repping
F. van der Veen

Human Reproduction Update 2010;16:577-89
Chapter 3

ABSTRACT

Background Various models have been developed for the prediction of pregnancy after IVF. These models differ from one another in the predictors they include. We performed a systematic review and meta-analysis to identify the most relevant predictors for success in IVF.

Methods We systematically searched MEDLINE and EMBASE for studies evaluating IVF/ICSI outcome. Studies were included if they reported an unconditional odds ratio or whenever one could be calculated for one or more of the following factors: age, type of infertility, indication, duration of infertility, basal FSH, number of oocytes, fertilization method, number of embryos transferred and embryo quality.

Results Fourteen studies were identified. A summary OR could be calculated for 5 factors. We found negative associations between pregnancy and female age (OR: 0.95, 95% CI: 0.94 to 0.96), duration of subfertility (OR: 0.99, 95% CI: 0.98 to 1.00) and basal FSH (OR: 0.94, 95% CI: 0.88 to 1.00). We found a positive association with number of oocytes (OR 1.04, 95% CI: 1.02 to 1.07). Better embryo quality was associated with higher pregnancy chances. No significant association was found for type of infertility and fertilization method. A summary OR for IVF indication and number of embryos transferred could not be calculated, because studies reporting on these used different reference categories.

Conclusions Female age, duration of subfertility, bFSH and number of oocytes, all reflecting ovarian function, are predictors of pregnancy after IVF. Better quality studies are necessary, especially studies that focus on embryo factors that are predictive of success in IVF.
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

INTRODUCTION

The first birth after in vitro fertilization (IVF) and embryo transfer was reported in 1978 (1). Initially, IVF was used to bypass infertility in women with bilateral tubal occlusion (2). In later years, IVF was also initiated in couples with unexplained subfertility, male subfertility, cervical factor, failed ovulation induction, endometriosis, or unilateral tubal pathology (3-5). In contrast to women with bilateral tubal occlusion, these women are not completely sterile but still have a chance of natural conception. To prevent overtreatment in these women it is important to balance the probability of achieving a pregnancy after IVF against the probability of achieving a pregnancy through natural conception.

Several cohort studies have identified factors that are possibly predictive of success after IVF, such as the diagnosis after the fertility workup, the number of previous unsuccessful IVF attempts and a previous pregnancy, with and without IVF (6-12). A useful prediction model for IVF success should include all relevant predictive factors, if these are available at a reasonable cost. Unfortunately, the putative predictive factors identified by these studies varied per study, and not all studies arrived at similar conclusions about factors predictive of IVF success.

To answer the question which factors can help in predicting pregnancy after IVF and should be included in an IVF prediction model, we performed a systematic review of the factors female age, parity, basal FSH, duration of subfertility, indication for subfertility, number of oocytes retrieved, method of fertilization, number of embryos transferred, and embryo quality to predict pregnancy after IVF, and to obtain pooled estimates of their predictive value through meta-analysis. These nine putative factors were chosen since they are routinely obtained in daily practice as part of standard patient care.

MATERIAL AND METHODS

Criteria for considering studies for this review
Articles were eligible if they evaluated the association between one or more of the pre-identified predictive factors and pregnancy after IVF/ICSI treatment in an unselected patient group. Articles were selected if the target population were subfertile women undergoing ovarian stimulation with gonadotrophins in fresh autologous IVF and ICSI procedures. The outcome measures were clinical pregnancy, defined as gestational sac confirmed by ultrasound at 6 weeks gestation, and ongoing pregnancy, defined as a pregnancy with heartbeat of one or more fetuses confirmed by ultrasound at 12 weeks gestation.
Chapter 3

Search strategy for the identification of studies
The searches were performed by a medical librarian (J.L.) experienced in conducting searches for systematic reviews. Literature searches were conducted in the bibliographic databases OVID MEDLINE and OVID EMBASE, from 1978 till August 2009, using both free-text words and index terms specific to each database (MeSH, SH). No language or any other restriction was applied. The search included an iterative process to refine the search strategy through adding search terms as new relevant citations were identified. We downloaded all references identified into Reference Manager® software (version 11.0).

To safeguard against missing relevant studies, we did not search for each of the nine individual factors separately (which might not be mentioned as such in title and abstract), but we searched for all prognostic studies on IVF or ICSI, using the following approach. A broad search for IVF/ICSI was combined with terms for pregnancy or pregnancy outcome (i.e. live birth). Next, this search was combined with two filters: [1] a broad search filter for prognostic methodology (based on terms as regression analysis, logistic models, multivariate or univariate or odds) and, separately [2], a broad search filter for prognostic/predictive factors (i.e. prognostic factor*, predictive factor*, independent* variable*). To check whether this search captured all relevant articles, we run a separate search for three individual factors (female age, basal FSH, number of embryos) without the abovementioned filters. This yielded no additional relevant articles. For details of the MEDLINE and EMBASE search see appendix I and II.

In and exclusion criteria
Articles were included if they reported on one or more studies that had evaluated associations between one or more predictive factors and pregnancy after IVF, if the study group consisted of subfertile women undergoing a fresh autologous IVF/ICSI cycle, and if a stimulation protocol with down regulation had been used.

Articles were excluded if they reported on a specific patient group within the subfertile IVF/ICSI population or if an unconditional odds ratio for the association between the putative predictive factor(s) and pregnancy was not reported and could not be calculated from the data presented.

Identification
The abstracts of all articles identified through the search were read by one researcher (L.L.), who selected all articles that were potentially eligible. In the next step, two researchers (L.L. and M.W.) carefully read and evaluated potentially eligible articles and decided on inclusion. In case of disagreement, the decision of a third reviewer (F.V.) was final. The reference list of every selected article was carefully checked to identify other potentially eligible studies.
Methods of review
The following information was extracted from each included article: study characteristics, (specified as consecutive or randomized study, prospectively or retrospectively, inclusion and exclusion criteria), predictors, outcome measures and their specific definitions (biochemical pregnancy defined as a positive pregnancy test, clinical pregnancy defined as ultrasonographic confirmation of an intrauterine gestation sac with fetal viability) and whether missing data were reported and/or imputed. If necessary, and whenever possible, we contacted the authors for missing data.

Statistical analyses
We extracted, calculated or recalculated the odds ratios for each predictor in each of the included articles, based on the data presented. We evaluated statistical heterogeneity graphically by drawing forest plots and by calculating the I² statistic. We then obtained summary estimates of the association by calculating the pooled unconditional odd ratio, using random effects modelling. The ORs of individual studies and summary odds ratios with corresponding confidence intervals were calculated using the Comprehensive Meta-Analysis software package (version 2).

RESULTS

Results of search
Our search retrieved 1,397 articles. The process of paper selection is summarized in Fig. 1. After screening titles and abstracts we selected 58 articles for further reading. Forty three articles did not meet our inclusion criteria, in particular in terms of reporting an unconditional odds ratio or allowing calculation of an odds ratio from the data presented (7;9;11-51). One article did not report on pregnancy or live birth as an outcome (52). A final number of 14 studies reporting on one or more of the predictive factors was included in the review.

Methodological quality of included studies
The characteristics of the 14 included studies are summarized in Table I. The number of evaluated predictors varied from 1 to 16. An overview of critical features of the included studies is shown in Fig. 2. Patient selection was consecutive in five (36%) studies. Only three studies (21%) had collected their data prospectively. Nine studies described their treatment protocol in sufficient detail. In 12 articles pregnancy was clearly defined. Only four studies reported on missing data. None of the studies used imputation for missing data.
Figure 1 | Process from initial search to final inclusion for papers on predictive factors in IVF/ICSI

Potentially relevant papers on predictive factors and IVF/ICSI
\( n = 1397 \)

Citations excluded after screening title and abstract
\( n = 1339 \)

Primary papers retrieved for full-text evaluation
\( n = 58 \)

Papers retrieved from cross references
\( n = 0 \)

Primary papers retrieved for full-text evaluation
\( n = 58 \)

Papers excluded after reading full-text paper
\( n = 44 \)

Reasons for exclusion:
- No unconditional odds ratio reported or could not be calculated  \( n = 43 \)
- No outcome of pregnancy or live birth  \( n = 1 \)

Primary papers included  \( n = 14 \)
- Age 13
- Primary or secondary subfertility 3
- Duration of subfertility 3
- Indication for IVF 4
- bFSH 7
- Number of oocytes retrieved 7
- Fertilization method 2
- Number of embryos transferred 2
- Embryo quality 3
Table I | Characteristics of the selected studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Patients</th>
<th>Inclusion and exclusion criteria</th>
<th>n</th>
<th>Study design</th>
<th>Outcome</th>
<th>Agonist/antagonist</th>
<th>Variables reported on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebbesen et al. (2009)</td>
<td>Women undergoing their first IVF-treatment cycle at a university fertility clinic</td>
<td>Inclusion: - First IVF cycle - No previous attempts with IVF-treatment - Ability to read and understand Danish Exclusion: - Preimplantation Genetic Diagnosis - Unplanned change of treatment type</td>
<td>837 ptn²</td>
<td>pros. CH¹</td>
<td>Clinical pregnancy</td>
<td>Agonist</td>
<td>Age, Smoking habits, Daily coffee, Stress measures, BMI, bFSH⁵, Method of fertilization, Number of oocytes</td>
</tr>
<tr>
<td>Sabatini et al. (2008)</td>
<td>Women undergoing their first IVF cycles</td>
<td>Inclusion: - Regular cycle in the previous 6 months Exclusion: - Woman’s age &gt; 45 years</td>
<td>1589 ptn</td>
<td>ret. CH</td>
<td>Live birth</td>
<td>Agonist</td>
<td>Age, bFSH</td>
</tr>
<tr>
<td>Wang et al. (2008)</td>
<td>Data from all fertility centres in Australia and New Zealand on women undergoing their first autologous fresh IVF/ICSI cycle</td>
<td>Inclusion: - Age woman ≥ 18 years - First autologous fresh cycle Exclusion: - Mixed fresh-thaw cycles - Gamete intrafallopian transfer cycles - Natural cycles - Surrogacy cycles</td>
<td>36412 ptn</td>
<td>ret. CH</td>
<td>Live birth/clinical pregnancy</td>
<td>NA⁴</td>
<td>Age</td>
</tr>
<tr>
<td>Ottosen et al. (2007)</td>
<td>IVF and ICSI treatment cycles from a public fertility clinic</td>
<td>Exclusion: - Cryo embryo transfer - Single embryo transfer</td>
<td>2193 cycl</td>
<td>ret. CH</td>
<td>Clinical pregnancy</td>
<td>Agonist or antagonist</td>
<td>Age, Duration of infertility, BMI, bFSH, Indication for IVF, Method of fertilization, Number of oocytes, Number of fertilized oocytes, Fertilization rate, Score of best/second best embryo</td>
</tr>
</tbody>
</table>
Continuation of table I

<table>
<thead>
<tr>
<th>Author</th>
<th>Patients</th>
<th>Inclusion and exclusion criteria</th>
<th>n</th>
<th>Study design</th>
<th>Outcome</th>
<th>Agonist/antagonist</th>
<th>Variables reported on</th>
</tr>
</thead>
</table>
| Ferlitsch et al. (2004) | Women referred for IVF to a university hospital | Inclusion:  
  - Weight and height known  
  - severe endometriosis  
  - a single ovary with possible normal ovarian response  
  - any ovarian cyst measuring > 10 mm in diameter on baseline day  
Exclusion:  
  - Weight and height known  
  - severe endometriosis  
  - a single ovary with possible normal ovarian response  
  - any ovarian cyst measuring > 10 mm in diameter on baseline day | 171 pt | ret. CH      | Clinical pregnancy | Agonist or antagonist | BMI  
LH  
E2  
Prolactin  
TSH  
Endometrium thickness  
Protocol |
| Hauzmaen et al. (2004) | Women who conceived after IVF/ICSI | Inclusion:  
  - Frozen archived serum sample for inhibin A measurement  
  - Only first pregnancy | 151 pt | ret. CH      | Ongoing/clinical pregnancy | Agonist | Age  
Number of oocytes  
Number of embryos transferred  
Day 11 hCG, level  
Mean inhibin A level |
| Huaault et al. (2002) | Patients from a university hospital in their first IVF cycle | Inclusion:  
  - Transfer of two embryos  
Exclusion:  
  - ICSI treatment  
  - Oocyte donation  
  - Cryo preserved embryos | 642 pt | ret. CH      | Ongoing pregnancy   | Agonist | Age  
Duration of infertility  
Type of infertility  
Indication for IVF  
Total number of sperm cells  
Progressive motile sperm cells  
Estrogen level  
Number of preovulatory follicles  
Number of oocytes retrieved  
Proportion of oocytes fertilized  
Day of ET  
No of embryos suitable for transfer  
Stage development best and second best embryo  
Morphology score of the best and second best embryo |
| Sharma et al. (2002) | Women undergoing IVF at an academic fertility centre | Exclusion:  
  - Cryo embryo transfers  
  - ICSI treatment | 2056 pt | ret. CH      | Clinical pregnancy | Agonist | Age  
Number of oocytes  
Number of embryos transferred |
<table>
<thead>
<tr>
<th>Study Authors</th>
<th>Study Design</th>
<th>Inclusion and Exclusion Criteria</th>
<th>Outcome Measures</th>
<th>Sample Size</th>
<th>Treatment</th>
<th>Selection</th>
<th>Agonist</th>
<th>Predictive Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mauguey-Laulom et al. (2002)</td>
<td>Women undergoing IVF or ICSI</td>
<td>Exclusion: - Women age ≥ 38 years and FSH &gt;10 UI/ml</td>
<td>Ongoing pregnancy</td>
<td>144 ptn</td>
<td>pros. CH</td>
<td>Ongoing</td>
<td>Agonist</td>
<td>Age, Endometrium thickness, Endometrium morphology, Pulsatility index, Protodiastole notch, Sub- and intra endometrial vascular signals</td>
</tr>
<tr>
<td>Hart et al. (2001)</td>
<td>All women undergoing their first IVF or ICSI</td>
<td>Inclusion: - Fibroids ≤ 5 cm Exclusion: - Cryo embryo transfers - Donated oocytes</td>
<td>Biochemical pregnancy</td>
<td>434 ptn</td>
<td>pros. CC</td>
<td>Biochemical</td>
<td>Agonist</td>
<td>Age, bFSH, Number of ampoules FSH, Number of oocytes, Number of available embryos, Intramural fibroid ≤ 5 cm in size</td>
</tr>
<tr>
<td>Bansci et al. (2000)</td>
<td>Women undergoing their first stimulated IVF cycle at an academic fertility centre</td>
<td>Inclusion: - Regular menstrual cycle - bFSH level on day 1-4 Exclusion: - Endocrine disorder - Oocyte donation - Unstimulated cycles</td>
<td>Ongoing pregnancy</td>
<td>435</td>
<td>ret. CH</td>
<td>Ongoing</td>
<td>Agonist</td>
<td>Age, Type of infertility, Indication for IVF, Duration of infertility, bFSH</td>
</tr>
<tr>
<td>Strandell et al. (2000)</td>
<td>Women undergoing IVF/ICSI</td>
<td>Inclusion: - Transfers with two embryos Exclusion: - Woman’s age &gt; 40 years - Cryo embryo transfers</td>
<td>Birth</td>
<td>1441 ptn</td>
<td>ret. CH</td>
<td>Birth</td>
<td>Agonist</td>
<td>Age, Previous pregnancy, Previous childbirth, Indication for IVF, FSH initial daily dose, Duration of ovarian stimulation, FSH total dose, Number of oocytes, Number of fertilized oocytes, Proportion of fertilized oocytes, Day of embryo transfer, Number of good quality embryos available, Number of good quality embryos transferred, No of embryos suitable for freezing</td>
</tr>
</tbody>
</table>
### Inclusion and exclusion criteria

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Inclusion criteria</th>
<th>n</th>
<th>Study design</th>
<th>Outcome</th>
<th>Agonist/antagonist</th>
<th>Variables reported on</th>
</tr>
</thead>
</table>
| Syrop et al. (1999) | Women undergoing their first IVF cycle | - Complete data available from first treatment cycle following determination of day 3 FSH/estriadiol and ovarian volume  
- Ovarian volume was determined by 1 of 2 physicians  
- Both ovaries were sonographically visualized  
- FSH/estriadiol determinations performed by same laboratory | 261 ptn | ret. CH | Clinical pregnancy | Agonist | Age  
Smoking (current/former)  
bFSH  
E2  
Smallest ovarian size |

Exclusion:  
- Anovulatory patients

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Inclusion criteria</th>
<th>n</th>
<th>Study design</th>
<th>Outcome</th>
<th>Agonist/antagonist</th>
<th>Variables reported on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stolwijk et al. (1997)</td>
<td>Women undergoing their first IVF or donor treatment in an academic fertility centre</td>
<td>- Normal uterine cavity</td>
<td>277 ptn</td>
<td>ret. CH</td>
<td>Ongoing/clinical/biochemical</td>
<td>Agonist</td>
<td>Age</td>
</tr>
</tbody>
</table>

Exclusion:  
- ICSI treatment  
- When there was no male partner

---

1 ICSI = intracytoplasmatic sperm injection  
2 ptn = patients; cycl = cycles  
3 Study design: pros. CH = prospective cohort study; pros. CC = prospective case control study; ret CH = retrospective cohort study  
4 NA = information not available  
5 bFSH = basal FSH  
6 E2 = estradiol
Predictor: age

Thirteen studies evaluated the association between female age and pregnancy after IVF (53-65). The characteristics of these studies are listed in Table I. The number of included patients varied from 144 to 36,412.

Three studies categorised age and data from these studies could not be pooled. One of these studies dichotomised age in two categories, ≤ 35 or > 35 years (61). Women aged 35 years or older had significantly lower pregnancy chances compared to women who were younger than 35 years. The second study categorized the patients into four categories, i.e. <30, 30-34, 35-38 and 39-45 years (60). Women in the age categories <30 and 30-34 years had 3.2 and 2.8 higher chances of a pregnancy compared to women in the age category 39-45 years. The third study showed that women aged 30 years or older compared to women in the age category 25 to 29 had lower pregnancy chances (65).

Figure 2 | Summary of study quality. Numbers indicate the number of studies.

<table>
<thead>
<tr>
<th>Predictor:</th>
<th>yes</th>
<th>no</th>
<th>unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consecutive patient selection</td>
<td>5</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Prospective data collection</td>
<td>3</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Protocol of treatment described in sufficient detail</td>
<td>9</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Pregnancy described in sufficient detail</td>
<td>12</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Missing data reported</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Missing data imputed</td>
<td>14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 | Forest plot presenting the effect of age on pregnancy after IVF/ICSI
Chapter 3

Age was reported as a continuous variable in the remaining 10 studies. Visual examination of the forest plot and the $I^2$ statistic (0%) suggested no heterogeneity across the studies (Fig. 3). The summary odds ratio for pregnancy and female age was 0.95 (95% CI: 0.94 to 0.96) indicating that increasing female age was associated with lower pregnancy chances in IVF.

**Predictor: duration of subfertility**

Three studies evaluated the association between duration of subfertility and pregnancy (53;57;59). One study subdivided duration of subfertility in six categories (59). The authors from that study reported that women with a duration of subfertility exceeding 12 months had lower pregnancy chances compared to women with a duration of subfertility of less than 12 months. In two studies duration of subfertility was taken as a continuous measurement and data could be pooled. Visual examination of the forest plot and the $I^2$ statistic (0%) suggested no heterogeneity across the studies (Fig. 4). The ongoing pregnancy rate per woman was lower with increasing duration of subfertility. The summary odds ratio of the two studies, reporting on 1,077 patients, was 0.99 (95% CI: 0.98 to 1.00).

**Figure 4** | Forest plot presenting the effect of duration of subfertility on pregnancy after IVF/ICSI

![Forest plot](attachment:image.png)

**Predictor: type of subfertility**

Three studies reported associations between type of subfertility (primary versus secondary subfertility) and pregnancy (53;57;63). One study reported that women with a previous clinical pregnancy had lower pregnancy chances after IVF, but women who previously had given birth had higher pregnancy chances after IVF. Neither of these associations was significant (63). Since this study did not report a 95% confidence interval, it could not be included in the meta analysis.

The data from two studies, including 1,077 cycles, could be pooled (53;57). Visual examination of the forest plot and the $I^2$ statistic (0%) suggested no heterogeneity between the studies (Fig. 5). The summary OR was 1.04 (95% CI: 0.65 to 1.43).
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

Figure 5 | Forest plot presenting the effect of type of subfertility on pregnancy after IVF/ICSI

![Forest plot showing effect of type of subfertility on pregnancy after IVF/ICSI]

Predictor: indication for IVF

Four studies reported the association between indication for IVF and pregnancy (53;57;59;63). One study evaluated this predictor using three categories: unexplained infertility, male infertility and tuboperitoneal disease. Unexplained infertility was used as the reference category. Women with male subfertility or tuboperitoneal disease had lower pregnancy chances compared to those with unexplained subinfertility (53). A second study reported that women with either male subfertility, tubal subfertility or subfertility caused by endometriosis had lower pregnancy chances compared to women with unexplained infertility (59).

In a third study the predictor “indication for IVF” was classified using four categories, with tubal subfertility as the reference category. Couples with male subfertility or with unexplained subfertility had lower pregnancy chances after IVF compared to couples with a tubal factor (57). The fourth study reported on each predictor separately. Women with tubal subfertility had significantly lower pregnancy chances after IVF and women with the indication endometriosis, male subfertility, unexplained subfertility and hormonal factors had higher pregnancy chances though not significant (63). Because of the use of different reference categories, we were not able to obtain a summary estimate of the odds ratio.

Predictor: basal FSH

Seven studies reported the association between basal FSH and pregnancy after IVF (53-55;59;60;64;66). Two of these studies (59;60) dichotomised basal FSH into the categories 0 to 10 IU and >10 IU. In both studies pregnancy chances were significantly higher in women with FSH <10 IU than in women with FSH concentrations of > 10 IU. The data of the remaining five studies could be pooled. The I² statistic (2%) suggested mild heterogeneity (Fig. 6). The summary OR confirmed that increasing bFSH values were associated with lower pregnancy rates after IVF (OR 0.94; 95% CI: 0.88 to 1.00).
Chapter 3

Figure 6 | Forest plot presenting the effect of basal FSH on pregnancy after IVF/ICSI

Predictor: number of oocytes retrieved

Six studies reported on the association between number of oocytes retrieved and pregnancy (55;57;59;61;63). Two studies had categorised the data. One study dichotomised number of oocytes in ≤ 5 and >5 oocytes retrieved (61). The other study used three categories: 1 to 5 oocytes, 6 to 10 and 11 or more oocytes (59). Both studies found that women with more oocytes had higher pregnancy chances.

The data of four studies could be pooled. Visual examination of the forest plot and the $I^2$ statistic (0%) suggested no heterogeneity across the studies (Fig. 7). We found a positive association between increasing number of oocytes retrieved and pregnancy chances after IVF, with a summary OR of 1.04 (95% CI: 1.02 to 1.07).

Figure 7 | Forest plot presenting the effect of number of oocytes retrieved on pregnancy after IVF/ICSI

Predictor: method of fertilization (IVF or ICSI)

Two studies reported on the association of method of fertilization and pregnancy chances after IVF (59;63). One study reported lower pregnancy chances with ICSI compared to IVF (OR 0.95, 95% CI: 0.79 to 1.14), though not significant (59). The other study showed no difference. This study did not report a 95% CI interval (63).
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

**Predictor: number of embryos transferred**
Two studies reported on the number of embryos transferred and IVF success (56;61). One study dichotomised the number into the categories more than 2 and 2 or less embryos transferred. Women where more than two embryos were transferred had significantly higher pregnancy chances (61). The second study showed higher, though not statistically significant, pregnancy chances when transferring more embryos (56). No summary OR could be calculated.

**Predictor: embryo quality**
Three studies evaluated the association between embryo quality and pregnancy after IVF (57;59;63). One study classified embryo quality using two separate factors, evaluating the best and the second best embryo in terms of stage of development and morphology score (57). The stage of development was described using 3 categories: delayed, appropriate and advanced stage. Advanced stage was used as the reference category. Women in whom either the best or second best embryo had a delayed or appropriate development stage had lower pregnancy chances compared to women where either the best or second best embryo had an advanced development stage. Lower morphology scores were also associated with lower pregnancy chances.

The second study reported that women with embryos with higher development stage and morphology scores, combined into one predictor, had higher pregnancy chances, compared to women with lower development stage and morphology score (59). The third study used three other predictors for embryo quality: number of good quality embryos available, number of good quality embryos transferred, and number of embryos suitable for freezing (63). All three predictors were associated with higher pregnancy chances after IVF. In all studies better embryo quality was associated with higher pregnancy chances, but, since these studies used different factors or combinations of embryo factors to report embryo quality, it was not possible to pool the data and calculate a summary OR.

**DISCUSSION**
Predicting pregnancy chances after an IVF cycle can help to prevent overtreatment and to balance the probability of achieving a pregnancy after IVF against the probability of achieving a pregnancy through natural conception. Although many studies reported on potential predictors of pregnancy chances after IVF, there is no consensus to pinpoint which predictors are clinically most relevant and on what factors one should base the decision to start treatment or not. In this systematic review and meta-analysis we evaluated nine putative predictive factors that could help in predicting pregnancy chances after IVF. Based on the available evidence we conclude that female age, duration of subfertility, basal FSH and number of oocytes are predictive of IVF success. Unfortunately we could not perform a meta-analysis on the factors indication for IVF, number of embryos transferred
and embryo quality, since there was no uniform method of reporting these variables. No meta-analysis was performed on the method of fertilization either, since only one study reported an OR and 95% confidence interval.

This meta-analysis provides robust evidence for female age being one of the strongest factors in predicting pregnancy chances after IVF. Our study not only shows that age is a significant predictor, it also shows that this predictor is identified by nearly every one of the included studies as an important predictor. So based on these findings, female age should not only be considered as a candidate predictor when developing a prognostic model for success in IVF, but the summary estimate from our meta-analysis could also be used as a prior estimate in a new prognostic model.

The biological explanation for the declining chances to conceive with increasing female age most likely lies in the diminished ovarian reserve, the decrease in both quantity and quality of oocytes, which is clinically relevant in women from their mid-30s. Diminished ovarian reserve generally leads to a poor response to gonadotropin therapy, and limits the possibility of a successful pregnancy. In our society many couples delay childbearing, which is illustrated by the mean age of women who become mothers for the first time; their age has increased over the last 17 years from 24.3 to 26.0 years.

The other factors we found to be associated with pregnancy chances, bFSH, duration of infertility and number of oocytes, are also age related. An older woman is likely to have a longer duration of subfertility, bFSH rises with increasing age and the number of oocytes declines with age. Unfortunately in this meta-analysis we were not able to perform a multivariable analysis and thus we do not known whether age in itself overrides these factors.

Although we could only include two studies reporting on the predictive value of number of embryos transferred and could not calculate a summary OR, there are several randomized controlled trials comparing fresh single embryo transfer to fresh double embryo transfer that clearly showed that double embryo transfer doubles the chance of pregnancy but also increases the risk of multiple pregnancy. These trials included ‘good prognosis’ women i.e. younger women without a history of multiple failed IVF cycles and with a certain number of good quality embryos available for transfer. However even in an unselected patient population the same results were found i.e. increased pregnancy chance but higher multiple pregnancy rate after double embryo transfer. The number of embryos transferred are thus not only predictive for pregnancy, but also for multiple pregnancy.

In addition to number of embryos, several studies have reported multivariable analyses that show that embryo quality in itself is a predictor of pregnancy chances in IVF, next to age. Our review shows that these studies did not use a uniform method
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

for reporting embryo quality. This made it impossible to perform a meta-analysis and to evaluate which embryo factor is most important. Since there are differences between studies on how they report embryo quality and differences in their selection criteria, it remains unclear which embryo factor is most predictive of pregnancy. Therefore studies on the relation between embryo quality and pregnancy need to use a standardised way of assessing embryo quality.

Several studies also showed that indication for IVF is a predictor for pregnancy (8;9;12). Since studies use different reference categories and different number of categories it was not possible to perform a meta-analysis. For future studies it would be useful to report every indication for IVF as a separate variable instead of combining all indications into one factor, to be able to compare all studies.

Our review of the literature on the nine predictors revealed that a remarkably few number of articles reported unconditional odds ratios, leaving only a few articles for inclusion. Maybe more data could be gathered, resulting in more precise summary estimates, in future IPD meta-analysis.

In summary, our systematic review shows that female age, duration of subfertility, basal FSH and number of oocytes are predictive for pregnancy chances after IVF. As a consequence these factors should be considered when making a decision to start treatment or not and the summary estimates could be used as a prior estimate in a new prognostic model. On the predictors indication for IVF, method of fertilization, number of embryos transferred and embryo quality we were not able perform a meta-analysis. Better quality studies are necessary, especially studies that focus on embryo factors that are predictive of success in IVF.
REFERENCES


Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis


22. Chuang CC, Chen CD, Chao KH, Chen SU, Ho HN, Yang YS. Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. Fertil Steril 2003 January;79(1):63-8.


31. Lee TH, Chen CD, Tsai YY, Chang LJ, Ho HN, Yang YS. Embryo quality is more important for younger women whereas age is more important for older women with regard to in vitro fertilization outcome and multiple pregnancy. Fertil Steril 2006 July;86(1):64-9.
Chapter 3

34. Lundin K, Bergh C, Hardarson T. Early embryo cleavage is a strong indicator of embryo quality in human IVF. Hum Reprod 2001 December;16(12):2652-7.
45. Terriou P, Sapin C, Giorgetti C, Hans E, Spach JL, Roulier R. Embryo score is a better predictor of pregnancy than the number of transferred embryos or female age.[see comment]. Fertil Steril 2001 March;75(3):525-31.
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis


Chapter 3

64. Syrop CH, Dawson JD, Husman KJ, Sparks AE, Van Voorhis BJ. Ovarian volume may predict assisted reproductive outcomes better than follicle stimulating hormone concentration on day 3. Hum Reprod 1999 July;14(7):1752-6.


77. Van Montfoort AP, Fiddelers AA, Janssen JM, Derhaag JG, Dirksen CD, Dunselman GA et al. In unselected patients, elective single embryo transfer prevents all multiples, but results in significantly lower pregnancy rates compared with double embryo transfer: a randomized controlled trial. Hum Reprod 2006 February;21(2):338-43.
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

Appendix I | Search Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R)

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fertilization in Vitro/ or Sperm Injections, Intracytoplasmic/</td>
</tr>
<tr>
<td>2</td>
<td>(Fertil?at* adj2 in Vitro).ti,ab.</td>
</tr>
<tr>
<td>3</td>
<td>(ivf or icsi or ic-si).ti,ab.</td>
</tr>
<tr>
<td>4</td>
<td>(intracytoplasm* adj5 sperm*).ti,ab.</td>
</tr>
<tr>
<td>5</td>
<td>embryo transfer/</td>
</tr>
<tr>
<td>6</td>
<td>(embryo adj2 (transfer* or implant*)).ti,ab.</td>
</tr>
<tr>
<td>7</td>
<td>reproductive techniques/ or reproductive techniques, assisted/</td>
</tr>
<tr>
<td>8</td>
<td>or/1-7</td>
</tr>
<tr>
<td>9</td>
<td>exp pregnancy/</td>
</tr>
<tr>
<td>10</td>
<td>Birth Rate/ or live birth/</td>
</tr>
<tr>
<td>11</td>
<td>exp Pregnancy Rate/</td>
</tr>
<tr>
<td>12</td>
<td>pregnan*.ti,ab.</td>
</tr>
<tr>
<td>13</td>
<td>(livebirth* or childbirth*).ti,ab.</td>
</tr>
<tr>
<td>14</td>
<td>((live or rate<em>1) adj1 birth</em>1).ti,ab.</td>
</tr>
<tr>
<td>15</td>
<td>(birth<em>1 adj2 (rate</em>1 or live or child*)).ti,ab.</td>
</tr>
<tr>
<td>16</td>
<td>((newborn<em>1 or child</em>) adj2 (live or born)).ti,ab.</td>
</tr>
<tr>
<td>17</td>
<td>(succes* adj (rate<em>1 or outcome</em>1)).ti,ab.</td>
</tr>
<tr>
<td>18</td>
<td>((gestation or reproducti<em>2) adj2 (achiev</em>2 or succes* or outcome*1)).ti,ab.</td>
</tr>
<tr>
<td>19</td>
<td>or/9-18</td>
</tr>
<tr>
<td>20</td>
<td>8 and 19</td>
</tr>
<tr>
<td>21</td>
<td>letter/ or case reports/ or randomized controlled trial/ or editorial/ or (animals/ not (animals/ and humans/))</td>
</tr>
<tr>
<td>22</td>
<td>20 not 21</td>
</tr>
<tr>
<td>23</td>
<td>regression analysis/ or logistic models/ or linear models/ or Multivariate Analysis/ or Odds ratio/</td>
</tr>
<tr>
<td>24</td>
<td>(multivaria* or univaria*).ti,ab.</td>
</tr>
<tr>
<td>25</td>
<td>odds.ti,ab.</td>
</tr>
<tr>
<td>26</td>
<td>(Crude adj (association* or relative risk*)).ti,ab.</td>
</tr>
<tr>
<td>27</td>
<td>((logistic* or log or coefficient) adj3 (regression* or model* or analys*)).ti,ab.</td>
</tr>
<tr>
<td>28</td>
<td>(multiple adj3 regression*).ti,ab.</td>
</tr>
<tr>
<td>29</td>
<td>(regression adj3 model*).ti,ab.</td>
</tr>
<tr>
<td>30</td>
<td>((linear or analys*) adj1 regression).ti,ab.</td>
</tr>
<tr>
<td>31</td>
<td>or/23-30</td>
</tr>
<tr>
<td>32</td>
<td>22 and 31</td>
</tr>
<tr>
<td>33</td>
<td>prognosticat*.ti,ab.</td>
</tr>
<tr>
<td>34</td>
<td>(prognos* adj2 value*).ti,ab. not (test* or assay* or imaging or ultrasonograph*).ti.</td>
</tr>
</tbody>
</table>
Chapter 3

Continuation of appendix I

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>(prognostic adj6 (factor* or variable or parameter* or indicator* or importance or index)).ti,ab.</td>
</tr>
<tr>
<td>36</td>
<td>(predict* adj2 factor*).ti,ab.</td>
</tr>
<tr>
<td>37</td>
<td>(predicting adj2 variable*).ti,ab.</td>
</tr>
<tr>
<td>38</td>
<td>(significa* or positive* or negat*) adj (predict? or predictor*).ti,ab.</td>
</tr>
<tr>
<td>39</td>
<td>(predict* adj3 (ongoing pregnan* or pregnancy rate*)).ti,ab.</td>
</tr>
<tr>
<td>40</td>
<td>(independent* or dependent*) adj4 (factor* or variable* or predictor* or associat* or marker*).ti,ab.</td>
</tr>
<tr>
<td>41</td>
<td>or/33-40</td>
</tr>
<tr>
<td>42</td>
<td>22 and 41</td>
</tr>
</tbody>
</table>

Appendix II | Search Embase

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>fertilization in vitro/ or intracytoplasmic sperm injection/</td>
</tr>
<tr>
<td>3</td>
<td>(ivf or icsi or ic-si).tw.</td>
</tr>
<tr>
<td>4</td>
<td>(intracytoplas* adj5 sperm*).tw.</td>
</tr>
<tr>
<td>5</td>
<td>embryo transfer/</td>
</tr>
<tr>
<td>6</td>
<td>(embryo adj2 (transfer* or implant*)).tw.</td>
</tr>
<tr>
<td>7</td>
<td>infertility therapy/</td>
</tr>
<tr>
<td>8</td>
<td>or/1-7</td>
</tr>
<tr>
<td>9</td>
<td>exp pregnancy/</td>
</tr>
<tr>
<td>10</td>
<td>birth rate/ or live birth/</td>
</tr>
<tr>
<td>11</td>
<td>pregnancy rate/</td>
</tr>
<tr>
<td>12</td>
<td>pregnan*.tw.</td>
</tr>
<tr>
<td>13</td>
<td>(livebirth* or childbirth*).tw.</td>
</tr>
<tr>
<td>14</td>
<td>(live or rate<em>1 adj1 birth</em>).tw.</td>
</tr>
<tr>
<td>15</td>
<td>(birth<em>1 adj2 (rate</em>1 or live or child*)).tw.</td>
</tr>
<tr>
<td>16</td>
<td>(newborn<em>1 or child</em>) adj2 (live or born)).tw.</td>
</tr>
<tr>
<td>17</td>
<td>(succes* adj (rate<em>1 or outcome</em>1)).tw.</td>
</tr>
<tr>
<td>18</td>
<td>((gestation or reproducti<em>2) adj2 (achiev</em>2 or succes* or outcome*1)).tw.</td>
</tr>
<tr>
<td>19</td>
<td>or/9-18</td>
</tr>
<tr>
<td>20</td>
<td>8 and 19</td>
</tr>
<tr>
<td>21</td>
<td>letter/ or case report/ or randomized controlled trial/ or editorial/ or ((animal or nonhuman) not human).hw.</td>
</tr>
<tr>
<td>22</td>
<td>20 not 21</td>
</tr>
</tbody>
</table>
Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis

Continuation of appendix II

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>regression analysis/ or linear regression analysis/ or logistic regression analysis/ or multiple linear regression analysis/ or multivariate logistic regression analysis/ or exp multivariate analysis/ or statistical model/</td>
</tr>
<tr>
<td>24</td>
<td>(multivaria* or univaria*).tw.</td>
</tr>
<tr>
<td>25</td>
<td>odds.tw.</td>
</tr>
<tr>
<td>26</td>
<td>(crude adj (association* or relative risk*)).tw.</td>
</tr>
<tr>
<td>27</td>
<td>((logistic* or log or coefficient) adj3 (regression* or model* or analy*)).tw.</td>
</tr>
<tr>
<td>28</td>
<td>(multiple adj3 regression).tw.</td>
</tr>
<tr>
<td>29</td>
<td>(regression adj3 model).tw.</td>
</tr>
<tr>
<td>30</td>
<td>((linear or analys*) adj1 regression).tw.</td>
</tr>
<tr>
<td>31</td>
<td>or/23-30</td>
</tr>
<tr>
<td>32</td>
<td>22 and 31</td>
</tr>
<tr>
<td>33</td>
<td>prognosticat*.tw.</td>
</tr>
<tr>
<td>34</td>
<td>(prognos* adj2 value*).tw. not (test* or assay* or imaging or ultrasonograph*).ti.</td>
</tr>
<tr>
<td>35</td>
<td>(prognostic adj6 (factor* or variable or parameter* or indicator* or importance or index)).tw.</td>
</tr>
<tr>
<td>36</td>
<td>(predict* adj2 factor*).tw.</td>
</tr>
<tr>
<td>37</td>
<td>(predicting adj2 variable*).tw.</td>
</tr>
<tr>
<td>38</td>
<td>((significa* or positive* or negat*) adj (predict? or predictor*)).tw.</td>
</tr>
<tr>
<td>39</td>
<td>(predict* adj3 (ongoing pregnanc* or pregnancy rate*)).tw.</td>
</tr>
<tr>
<td>40</td>
<td>((independent* or dependent*) adj4 (factor* or variable* or predictor* or associat* or marker*)).tw.</td>
</tr>
<tr>
<td>41</td>
<td>or/33-40</td>
</tr>
<tr>
<td>42</td>
<td>22 and 41</td>
</tr>
<tr>
<td>43</td>
<td>42 not 32</td>
</tr>
</tbody>
</table>
Individualized decision-making in IVF: calculating the chances of pregnancy

L.L. van Loendersloot
M. van Wely
S. Repping
P.M.M. Bossuyt
F. van der Veen

Human Reproduction 2013; 28:2972-80
Study question:
Are we able to develop a model to calculate the chances of pregnancy prior to the start of the first IVF cycle as well as after one or more failed cycles?

Summary answer:
Our prediction model enables the accurate individualized calculation of the probability of an ongoing pregnancy with IVF.

What is known already:
To improve counselling, patient selection and clinical decision-making in IVF, a number of prediction models have been developed. These models are of limited use as they were developed before current clinical and laboratory protocols were established.

Study design, size, duration:
This was a cohort study. The development set included 2621 cycles in 1326 couples who had been treated with IVF or ICSI between January 2001 and July 2009. The validation set included additional data from 515 cycles in 440 couples treated between August 2009 and April 2011. The outcome of interest was an ongoing pregnancy after transfer of fresh or frozen–thawed embryos from the same stimulated IVF cycle. If a couple became pregnant after an IVF/ICSI cycle, the follow-up was at a gestational age of at least 11 weeks.

Participants/materials, setting, methods:
Women treated with IVF or ICSI between January 2001 and April 2011 in a university hospital. IVF/ICSI cycles were excluded in the case of oocyte or embryo donation, surgically retrieved spermatozoa, patients positive for human immunodeficiency virus, modified natural IVF and cycles cancelled owing to poor ovarian stimulation, ovarian hyperstimulation syndrome or other unexpected medical or non-medical reasons.

Main results and the role of chance:
Thirteen variables were included in the final prediction model. For all cycles, these were female age, duration of subfertility, previous ongoing pregnancy, male subfertility, diminished ovarian reserve, endometriosis, basal FSH and number of failed IVF cycles. After the first cycle: fertilization, number of embryos, mean morphological score per Day 3 embryo, presence of 8-cell embryos on Day 3 and presence of morulae on Day 3 were also included. In validation, the model had moderate discriminative capacity (c-statistic 0.68, 95% confidence interval: 0.63–0.73) but calibrated well, with a range from 0.01 to 0.56 in calculated probabilities.

Limitations, reasons for caution:
In our study, the outcome of interest was ongoing pregnancy. Live birth may have been a more appropriate outcome, although only 1–2% of all ongoing pregnancies result in late miscarriage or stillbirth. The model was based on data from a single centre.
Individualized decision-making in IVF: calculating the chances of pregnancy.

**Wider implications of the findings:**
The IVF model presented here is the first to calculate the chances of an ongoing pregnancy with IVF, both for the first cycle and after any number of failed cycles. The generalizability of the model to other clinics has to be evaluated more extensively in future studies (geographical validation). Centres with higher or lower success rates could use the model, after recalibration, by adjusting the intercept to reflect the IVF success rates in their centre.

**Study funding/competing interest(s):**
This project was funded by the NutsOhra foundation (Grant 1004-179). The NutsOhra foundation had no role in the development of our study, in the collection, analysis and interpretation of data; in writing of the manuscript, and in the decision to submit the manuscript for publication. There were no competing interests.
Since the introduction of in vitro fertilization (IVF) in 1978 over 3.75 million babies have been born worldwide using IVF (1). IVF is currently one of the most widely used intervention for infertility. In 2007, 376,971 treatment cycles were reported in 18 European countries, 142,435 cycles in the USA and 56,817 cycles in Australia and New Zealand (2-4).

IVF is considered as a last resort for all infertile couples regardless of the aetiology of their infertility (5-7). In contrast to patients’ perceptions, IVF does not guarantee success; almost 50% of couples that start IVF will remain childless, even if they undergo multiple IVF cycles (8). Given this limited success, it seems logic to offer IVF only to couples with reasonable chances of success and to discontinue treatment when chances are low and no longer outweigh the burden and costs.

To improve counselling, patient selection and clinical decision making in IVF, a number of prediction models have been developed in the past (9). Several models are of limited use since they were developed before current clinical and laboratory protocols were established (10-20). Most models do not include the transfer of frozen–thawed embryos, an essential component of modern day IVF (10-27). In Europe alone almost 86,059 frozen-thawed transfers were performed in 2006 resulting in 10,382 pregnancies, -constituting approximately 15% of all pregnancies achieved in that year (28). A number of models calculate pregnancy chances only for the first IVF cycle, while others calculate pregnancy chances after one failed IVF cycle only (19;29). This limits their practical use, since the average pregnancy rate is approximately 29% per cycle, and thus in over 70% of the couples a decision has to be made whether or not to continue IVF (28). We therefore set out to develop a model that would calculate pregnancy chances prior to the start of the first IVF cycle as well as after one or more failed cycles. The model is based on empirical data systematically collected in consecutive IVF patients.

METHODS

Patients

We collected data in a historical cohort of couples that had been treated with IVF or ICSI between January 2001 and July 2009 in the Centre for Reproductive Medicine of the Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands. This will be referred to as the development set.

All couples in our cohort had been trying to conceive for at least 12 months. They had undergone a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (30). The indication to start IVF or ICSI treatment was
Individualized decision-making in IVF: calculating the chances of pregnancy.

determined according to the Dutch IVF guideline (5). If subfertility was caused by tubal pathology, such as two-sided tubal blockage and severe endometriosis, or severe oligozoospermia (post-wash total motile sperm count < 3 million) IVF/ICSI was offered directly (31). In the case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia, and unexplained subfertility, at least six intra uterine inseminations (IUI) were applied before IVF/ICSI was offered. In the case of ovulation disorders, mainly caused by polycystic ovary syndrome (PCOS), 12 cycles of ovulation induction were applied before IVF/ICSI was offered.

Data on clinical diagnoses, IVF protocol and response, and laboratory data on embryo morphology and growth, as well as treatment outcomes for all IVF/ICSI cycles were retrieved from our clinical databases and medical records. Included in the analyses were data on stimulated IVF/ICSI cycles and also from frozen–thawed embryo transfers from these stimulated cycles. We excluded IVF/ICSI cycles that involved oocyte or embryo donation, cycles that used surgically retrieved spermatozoa, cycles from human immunodeficiency virus-positive patients, cycles that involved a modified natural cycle and cycles cancelled due to poor ovarian stimulation, ovarian hyperstimulation syndrome or other unexpected medical or non-medical reasons (32). Women underwent controlled ovarian hyperstimulation after down-regulation with the GnRH agonist triptorelin (Decapeptyl®) in a long protocol with a midluteal start. Controlled ovarian hyperstimulation was started on cycle day 5 with recombinant FSH or HMG in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Follicular maturation was induced by 10,000 IU human chorionic gonadotropin hormone (hCG) (Pregnyl, Organon). Cumulus-oocyte complexes were recovered by transvaginal ultrasound guided follicle aspiration 36 hours thereafter. Oocytes were inseminated with 10,000 or 15,000 progressively motile spermatozoa (in vitro fertilization) or injected with a single spermatozoon (intracytoplasmic sperm injection) 2-4 hours after follicle aspiration. Embryos were cultured in Human Tubal Fluid (HTF, Cambrex) or G5 medium (Vitrolife) at 37°C and 5% CO₂ in air. Embryo transfer was performed mostly 72h and occasionally 96h after follicle aspiration with a Wallace catheter (Smiths Medical). Supernumerary embryos of good quality were frozen on day 4 after follicle aspiration using a slow-freeze protocol. Luteal phase was supported by progesterone intravaginally two times 200 mg (Utrogestan) per day. A hCG blood test was performed 18 days after oocyte retrieval.

Each embryo was cultured individually. On each day of development the number of blastomeres was assessed and each embryo was given a morphological score. For the morphological score the degree of fragmentation of the embryo and the uniformity of the blastomeres were assessed (33). Based on the degree of fragmentation embryos were scored as 1 (no fragments), 2 (<20% fragmentation), 3 (20-50% fragmentation) or 4 (> 50% fragmentation). If the blastomeres of the embryo were non-uniform in size the morphological score was reduced with one point with 4 remaining the lowest possible
score. If on day 3 the embryo showed signs of compaction the embryo was scored as a morula and given a grade based on the degree of compaction (score 1: full compaction, score 2: >0-20% compaction and score 3: < 20% compaction).

Outcome
The outcome of interest was an ongoing pregnancy after transfer of fresh or frozen-thawed embryos from the same stimulated IVF cycle. Ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal ultrasound at a gestational age of at least 11 weeks.

Data analysis
We first analysed our data with generalized estimating equations and afterward with logistic regression. The point estimates and confidence intervals after analysis with generalized estimation equations were almost identical to those of logistic regression. As logistic models are easier to interpret and the point estimates did not differ, we decide to use a multivariable logistic regression to develop a model.

A model was developed to calculate the probability of pregnancy after IVF, including fresh and frozen-thawed embryo transfers from the same cycle. We identified a number of candidate predictors based on a recent systematic review and meta-analysis and on different cohort studies on predictive factors in IVF, reported elsewhere (20;22;34;35). The list of candidate predictors included clinical characteristics, available before the start of IVF (female and male age, previous pregnancies, duration of subfertility, indication for IVF, and basal FSH), IVF stimulation parameters (initial FSH dose), and laboratory data from the previous failed IVF cycle, if applicable [fertilization method (IVF/ICSI), number of oocytes, number of embryos, embryo quality and number of embryos transferred].

Some of the candidate predictors had missing values. Simple exclusion of couples with missing values on one or more variables commonly causes biased results and decreases statistical efficiency (36). For this reason, we first performed an analysis with missingness indicators and then completed missing values by multiple imputation using Statistical Package for the Social Sciences (version 18.0) (36). This method uses all available data to impute the missing values based on the correlation between each variable with missing values and all other variables. If candidate predictors had 25% or more missing values, they were excluded from the analyses.

We first checked the linearity of the association between the continuous variables female age, male age, duration of subfertility, basal FSH, initial FSH dose, number of embryos and the logit transformed probability of an ongoing pregnancy using restricted cubic spline functions in univariable logistic regression. We performed similar preliminary analyses for the number of blastomeres, morphological score and embryo implantation. The analysis demonstrated a nonlinear association between the continuous variables female age, male
Individualized decision-making in IVF: calculating the chances of pregnancy.

age, duration of subfertility, basal FSH, initial FSH dose, number of embryos and ongoing pregnancy after IVF. We therefore transformed all variables to better fit the data. Age was transformed using a polynomial: $\text{Age} + \text{Age}^2 + \text{Age}^3$. The duration of subfertility was capped at 5 years. Basal FSH was capped at the bottom with values below 10 U/l set at 10 U/l. Initial dose FSH was similarly capped with values above 300 IU coded as 300 IU. The number of embryos was capped at 10 embryos. All embryo morphological scores could adequately be described using linear functions. The number of blastomeres on day 2 was recoded as the absolute value of the deviation from 4. The number of blastomeres on day 3 was recoded as the absolute value of the deviation from 8.

For each candidate predictor, we performed a univariable logistic regression analysis and estimated the corresponding unconditional odds ratio, 95% confidence intervals (CI), and P-value.

Since we wanted to obtain a model that would rely, as much as possible, on parsimonious data collection, we used a blockwise model building strategy. We started with data available before the initiation of IVF. We were only prepared to add data from previous failed cycles and laboratory parameters if they sufficiently contributed to model fit. We therefore started our model building with the patient characteristics. All features that were associated with ongoing pregnancy were entered in a multivariable logistic regression analysis. For reasons of parsimony, we removed variables from the model if their removal did not significantly reduce model fit, using the generalized likelihood ratio test statistics.

In the next step, we considered embryo characteristics and used a strategy similar to that employed for the patient characteristics, first adding all embryo characteristics associated with ongoing pregnancy and then removing redundant embryo characteristics one by one, based on the generalized likelihood ratio test statistic. In a third and final step, we used a similar approach for the IVF stimulation parameters.

We explicitly tested whether a model with different point estimates for each parameter depending on the cycle number had a better fit than a simpler model using cycle number as a parameter and similar point estimates, regardless of the cycle number, for each parameter. If both models showed similar results we continued using the simpler model.

As the capacity to predict ongoing pregnancy of a variable may vary in a series of IVF cycles, we explicitly tested statistically for interactions between included predictors and IVF cycle number. In deciding between competing expressions of related parameters, we used Akaike’s Information Criterion (AIC) in variable selection.

To prevent overfitting and to avoid a too optimistic impression of model performance, a linear shrinkage factor was estimated based on model fit and the number of parameters (37). Coefficients in the model were then corrected by this shrinkage factor.
Chapter 4

Performance
The performance of the model was first evaluated by assessing the ability of the model to distinguish between women who achieved ongoing pregnancy and those who did not (discrimination). We calculated the area under the receiver operating characteristic curve (AUC), also known as the c-statistic.

To evaluate agreement between calculated probabilities of an IVF pregnancy and observed proportions of achieving a pregnancy, we performed the Hosmer and Lemeshow goodness-of-fit test statistic. In addition, we compared the average calculated probabilities of an ongoing pregnancy in disjoint subgroups defined by quintiles with the observed ongoing pregnancy rate in the corresponding groups in a calibration plot.

To evaluate any miscalibration, we also fitted a calibration model using logistic regression, with the linear combination of variables in the prediction model as the only variable (37;38).

External validation
A prediction model may not perform as well in new patients as in the development set (37). We performed an external, temporal validation using more recent data, collected at the same clinic after the data used for the development of the model (37). We validated our model on data of all couples who had been treated with IVF/ICSI from August 2009 till April 2011 in the Centre for Reproductive Medicine of the Academic Medical Centre, the Netherlands.

Updating the model
To obtain a model with better precision and stronger validity, we updated the coefficients the final model after the external validation by re-calibration (39;40).

RESULTS

We could include data from 1,326 couples who had undergone 2,621 cycles: of which 1,421 were first IVF cycles, 729 were second IVF cycles, 339 were third IVF cycles and 132 were four up to eighth cycles. Two thousand one hundred ninety-six fresh embryo transfers were conducted 72 hours after oocyte retrieval, 202 fresh embryo transfers were conducted 96 hours after oocyte retrieval and in 223 cycles, there was no suitable embryo for transfer. There was a total of 903 frozen-thawed cycles, 549 after the first IVF cycle, 229 after the second IVF cycle 104 after the third IVF cycle and 21 after the fourth till eighth IVF cycle. There were 570 ongoing pregnancies from fresh transfers and 82 ongoing pregnancies from frozen-thawed embryo transfer, yielding a total of 652 ongoing pregnancies (24.9% per cycle). The baseline characteristics of the couples are summarized in Table I.
Individualized decision-making in IVF: calculating the chances of pregnancy.

Table I Baseline characteristics of the cycles included in the development and validation datasets

<table>
<thead>
<tr>
<th></th>
<th>Development set (n=2621)</th>
<th>Validation set (n=515)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female age years (SD)</td>
<td>35.3 (4.6)</td>
<td>36.71 (4.9)</td>
</tr>
<tr>
<td>Male age years (SD)</td>
<td>38.4 (6.6)</td>
<td>39.9 (6.7)</td>
</tr>
<tr>
<td>Duration of subfertility (years)</td>
<td>3.8 (2.4)</td>
<td>3.9 (2.5)</td>
</tr>
<tr>
<td><strong>Type of subfertility</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary subfertility (%)</td>
<td>1860 (71%)</td>
<td>346 (67%)</td>
</tr>
<tr>
<td>Secondary subfertility (%)</td>
<td>761 (29%)</td>
<td>169 (33%)</td>
</tr>
<tr>
<td>Previous ongoing pregnancy</td>
<td>569 (22%)</td>
<td>138 (27%)</td>
</tr>
<tr>
<td>FSH (SD)</td>
<td>7.7 (3.6)</td>
<td>7.9 (4.0)</td>
</tr>
<tr>
<td>BMI (SD)</td>
<td>24.3 (4.7)</td>
<td>25.1 (5.5)</td>
</tr>
<tr>
<td><strong>Indication for IVF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexplained subfertility (%)</td>
<td>518 (20%)</td>
<td>141 (27%)</td>
</tr>
<tr>
<td>Tubal pathology (%)</td>
<td>587 (22%)</td>
<td>69 (13%)</td>
</tr>
<tr>
<td>Male subfertility (%)</td>
<td>1352 (52%)</td>
<td>227 (44%)</td>
</tr>
<tr>
<td>Polycystic ovary syndrome (%)</td>
<td>217 (8%)</td>
<td>18 (3%)</td>
</tr>
<tr>
<td>Diminished ovarian reserve (%)</td>
<td>215 (8%)</td>
<td>80 (16%)</td>
</tr>
<tr>
<td>Endometriosis (%)</td>
<td>118 (5%)</td>
<td>25 (5%)</td>
</tr>
<tr>
<td>Cervical hostility (%)</td>
<td>77 (3%)</td>
<td>10 (2%)</td>
</tr>
<tr>
<td><strong>Number of previous failed IVF/ICSI cycles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 failed IVF/ICSI cycle (%)</td>
<td>1421 (54%)</td>
<td>312 (61%)</td>
</tr>
<tr>
<td>1 failed IVF/ICSI cycle (%)</td>
<td>729 (28%)</td>
<td>121 (23%)</td>
</tr>
<tr>
<td>2 failed IVF/ICSI cycles (%)</td>
<td>339 (13%)</td>
<td>62 (12%)</td>
</tr>
<tr>
<td>3-7 failed IVF/ICSI cycles (%)</td>
<td>132 (5%)</td>
<td>20 (4%)</td>
</tr>
<tr>
<td><strong>IVF stimulation parameters of the previous failed IVF/ICSI cycle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSH initial dose (mean, SD)</td>
<td>208 (112)</td>
<td>287 (133)</td>
</tr>
<tr>
<td><strong>Type of fertilization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVF (%)</td>
<td>1338 (51%)</td>
<td>247 (48%)</td>
</tr>
<tr>
<td>ICSI (%)</td>
<td>1239 (47%)</td>
<td>268 (52%)</td>
</tr>
<tr>
<td><strong>Embryological data of the previous failed IVF/ICSI cycle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of oocytes (SD)</td>
<td>9.31 (5.6)</td>
<td>9.43 (5.1)</td>
</tr>
<tr>
<td>Normal fertilization % (SD)</td>
<td>53.8 (29.3)</td>
<td>45.6 (27.3)</td>
</tr>
<tr>
<td>Number of embryos (SD)</td>
<td>5.04 (4.1)</td>
<td>4.54 (3.8)</td>
</tr>
<tr>
<td>Mean no. of cells per embryo on day 3 (SD)</td>
<td>5.33 (2.3)</td>
<td>4.68 (2.2)</td>
</tr>
<tr>
<td>No. of eight-cell embryos on day 3 (SD)</td>
<td>1.15 (1.7)</td>
<td>0.61 (1.0)</td>
</tr>
<tr>
<td>No. of morulae on day 3(SD)</td>
<td>0.09 (0.4)</td>
<td>0.04 (0.3)</td>
</tr>
</tbody>
</table>
Chapter 4

Continuation of table I

<table>
<thead>
<tr>
<th></th>
<th>Development set (n=2621)</th>
<th>Validation set (n=515)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of embryos with optimal progression (day 2 four-cell and day 3 eight cells, SD)</td>
<td>0.90 (1.5)</td>
<td>0.39 (0.9)</td>
</tr>
<tr>
<td>Mean morphological score, all embryos day 3 (SD)</td>
<td>2.2 (0.8)</td>
<td>2.2 (0.9)</td>
</tr>
<tr>
<td>Number of frozen embryos (SD)</td>
<td>1.12 (2.4)</td>
<td>0.49 (1.1)</td>
</tr>
<tr>
<td>Number of embryos transferred (SD)</td>
<td>1.70 (0.7)</td>
<td>1.86 (1.0)</td>
</tr>
</tbody>
</table>

Two variables had missing values, i.e. duration of subfertility (< 0.001% missing) and basal FSH (18% missing). The missingness indicator variables were not significant in the analysis described below.

Univariable analysis confirmed that younger women, younger men, couples with a shorter duration of subfertility, with secondary subfertility instead of primary subfertility, those having achieved a previous ongoing pregnancy, those with lower basal FSH, a diagnosis of male subfertility, a diagnosis of PCOS, lower initial dose of FSH, more oocytes, more embryos, more morulae on day 3 and more frozen embryos had significantly higher chances of an ongoing pregnancy with IVF. A diagnosis of diminished ovarian reserve, a diagnosis of endometriosis, and more failed IVF cycles were significantly associated with lower chances of an ongoing pregnancy.

Thirteen predictors were included in the final multivariable logistic regression model. These were the following patient characteristics: female age, duration of subfertility, previous ongoing pregnancy, male subfertility, diminished ovarian reserve, endometriosis, basal FSH, and number of failed IVF cycles. We added an interaction term for female age and male subfertility, and one for diminished ovarian reserve and endometriosis. For the calculation of pregnancy chances after the first cycle we added the following embryo features from the previous cycle (if any) to the patient characteristics: fertilization, number of embryos, mean morphological score per day 3 embryo, presence of eight-cell embryos on day 3, and presence of morulae on day 3 (Table II).

There was no significant additional effect of IVF cycle number, nor were there any significant interactions between the identified predictors and cycle number. For this reason we used the same point estimates for all predictor and included cycle number as a predictor (Suppl Table).
Individualized decision-making in IVF: calculating the chances of pregnancy.

Table II | Multivariable analysis for predicting pregnancy chances after an IVF/ICSI cycle

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Original model</th>
<th></th>
<th></th>
<th>Updated model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β*</td>
<td>95% CI</td>
<td>P-value</td>
<td>β</td>
<td>95% CI</td>
</tr>
<tr>
<td>Intercept</td>
<td>26.0950</td>
<td></td>
<td></td>
<td>22.0109</td>
<td></td>
</tr>
<tr>
<td><em>Patient characteristics</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-2.5792</td>
<td>(-4.76 - -0.40)</td>
<td>0.01</td>
<td>-2.1909</td>
<td>(-4.37 - -0.01)</td>
</tr>
<tr>
<td>Age²</td>
<td>0.0851</td>
<td>(0.02 - 0.15)</td>
<td>0.01</td>
<td>0.0723</td>
<td>(0.01 - 0.14)</td>
</tr>
<tr>
<td>Age³</td>
<td>-0.0009</td>
<td>(0.00 - 0.00)</td>
<td>0.00</td>
<td>-0.0008</td>
<td>(0.00 - 0.00)</td>
</tr>
<tr>
<td>Duration of subfertility†</td>
<td>-0.1001</td>
<td>(-0.18 - -0.02)</td>
<td>0.00</td>
<td>-0.0850</td>
<td>(-0.16 - -0.01)</td>
</tr>
<tr>
<td>Previous ongoing pregnancy</td>
<td>0.2338</td>
<td>(0.00 - 0.47)</td>
<td>0.03</td>
<td>0.1986</td>
<td>(-0.04 - 0.43)</td>
</tr>
<tr>
<td>Male subfertility</td>
<td>1.0880</td>
<td>(-0.55 - 2.72)</td>
<td>0.15</td>
<td>0.9242</td>
<td>(-0.71 - 2.56)</td>
</tr>
<tr>
<td>Diminished ovarian reserve</td>
<td>-0.9239</td>
<td>(-1.50 - -0.35)</td>
<td>0.00</td>
<td>-0.7848</td>
<td>(-1.36 - -0.21)</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>-0.5635</td>
<td>(-1.11 - -0.02)</td>
<td>0.03</td>
<td>-0.4786</td>
<td>(-1.02 - 0.07)</td>
</tr>
<tr>
<td>Basal FSH‡</td>
<td>-0.0798</td>
<td>(-0.16 - 0.00)</td>
<td>0.04</td>
<td>-0.0678</td>
<td>(-0.15 - 0.01)</td>
</tr>
<tr>
<td>Number of previous failed IVF cycles</td>
<td>-0.2391</td>
<td>(-0.43 - -0.05)</td>
<td>0.01</td>
<td>-0.2031</td>
<td>(-0.40 - -0.01)</td>
</tr>
<tr>
<td>Age * male subfertility</td>
<td>-0.0322</td>
<td>(-0.08 - -0.01)</td>
<td>0.14</td>
<td>-0.0274</td>
<td>(-0.07 - 0.02)</td>
</tr>
<tr>
<td>Endometriosis * diminished ovarian reserve</td>
<td>1.7872</td>
<td>(0.25 - 3.32)</td>
<td>0.01</td>
<td>1.5181</td>
<td>(-0.02 - 3.05)</td>
</tr>
<tr>
<td><em>Embryo parameters</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embryo yes/no after ovum pick up</td>
<td>0.8503</td>
<td>(-0.02 - 1.72)</td>
<td>0.04</td>
<td>0.7223</td>
<td>(-0.15 - 1.59)</td>
</tr>
<tr>
<td>Number of embryos after ovum pick up¥</td>
<td>0.0610</td>
<td>(0.00 - 0.12)</td>
<td>0.02</td>
<td>0.0518</td>
<td>(0.00 - 0.11)</td>
</tr>
<tr>
<td>Mean morphological score all embryos day 3</td>
<td>-0.3613</td>
<td>(-0.69 - -0.03)</td>
<td>0.02</td>
<td>-0.3069</td>
<td>(-0.64 - 0.02)</td>
</tr>
<tr>
<td>Eight-cell embryo yes/no on day 3</td>
<td>-0.3315</td>
<td>(-0.66 - 0.00)</td>
<td>0.03</td>
<td>-0.2816</td>
<td>(-0.61 - 0.05)</td>
</tr>
<tr>
<td>Morulae yes/no on day 3</td>
<td>0.6219</td>
<td>(0.09 - 1.15)</td>
<td>0.01</td>
<td>0.5283</td>
<td>(0.00 - 1.06)</td>
</tr>
</tbody>
</table>

Age² = Age squared
Age³ = Age to the power of 3
† duration of subfertility ≥ 5 years = 5 years
‡ basal FSH (bFSH) ≤ 10 IE/L = 10 IE/L
¥ number of embryos ≥ 10 = 10 embryos
β* = corrected beta coefficient for overfit

The calculated probabilities of an ongoing pregnancy for the 1,326 couples in the development set had a wide range: from 0.00 to 0.72, with a mean of 0.25 (Figure 1). Twenty-five per cent of the cycles had a probability of a pregnancy of less than 0.17, 25% had a probability between 0.18 and 0.26, 25% a probability between 0.27 and 0.32, and 25% had a probability exceeding 0.33. Four hypothetical cases and the corresponding probabilities are shown as an example in Table III.
### Table III | Four hypothetical patients with their calculated ongoing pregnancy chance in their subsequent IVF/ICSI cycle

<table>
<thead>
<tr>
<th></th>
<th>Patient A</th>
<th>Patient B</th>
<th>Patient C</th>
<th>Patient D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>34</td>
<td>42</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td><strong>Pregnancy history</strong></td>
<td>None</td>
<td>None</td>
<td>Miscarriage not after IVF</td>
<td>None</td>
</tr>
<tr>
<td><strong>Cause of infertility</strong></td>
<td>Unexplained subfertility</td>
<td>Male subfertility and diminished ovarian reserve</td>
<td>Diminished ovarian reserve</td>
<td>Male subfertility</td>
</tr>
<tr>
<td><strong>Duration of infertility</strong></td>
<td>4 years</td>
<td>4 years</td>
<td>7 years</td>
<td>2 years</td>
</tr>
<tr>
<td><strong>Previous IVF cycles</strong></td>
<td>Two</td>
<td>One</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td><strong>Data from last IVF cycle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of embryos after ovum pick up</strong></td>
<td>7</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mean morphological score all embryos day 3</strong></td>
<td>2.0</td>
<td>2.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Eight-cell embryo yes/no on day 3</strong></td>
<td>yes</td>
<td>no</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Morulae yes/no on day 3</strong></td>
<td>no</td>
<td>no</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Calculated probability of an ongoing pregnancy</strong></td>
<td>0.25</td>
<td>0.05</td>
<td>0.13</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Individualized decision-making in IVF: calculating the chances of pregnancy.

The model had moderate discriminative capacity in the development set. The c-statistic was 0.68 (95% CI: 0.65 to 0.70). In the development set the model calibrated well; the goodness-of-fit test (Hosmer-Lemeshow) showed no significant miscalibration (p=0.41). Figure 2 shows the calibration plot. In case of perfect calibration, all points would be on the diagonal, the line of equality, and average probabilities correspond perfectly to the observed pregnancy rates. Our calibration plot showed that the model calibrated well (Figure 2). In the calibration model, the estimated intercept was 0.10 (95% CI: -0.10 to 0.29) and the slope 1.10 (95% CI: 0.92 to 1.27). This intercept reflects the extent to which predictions are systematically too low or too high, i.e. ‘calibration-in-the-large’. Ideally, the intercept is zero and the slope unity.

Figure 2 | Calibration plots, showing the association between the calculated and observed rates of ongoing pregnancy after IVF/ICSI

The validation was performed on data from 440 couples undergoing 515 cycles of IVF. Baseline characteristics are summarized in Table I. The calculated probabilities of an ongoing pregnancy for the 515 cycles in the validation set ranged from 0.01 to 0.56, with a mean of 0.22, indicative of a population with more cycles with intermediate and poor prognosis compared with the cycles in the development set (Figure 1).

The discriminative capacity was similar to that in the development set, with a c-statistic of 0.68 (95% CI: 0.63 to 0.73). The model calibrated well for the first three quintiles, with calculated probabilities in the range from 0.0 to 0.26. The model somewhat underestimated the actual rate in the fourth quintile (calculated probability in the range from 0.26 to 0.32) and overestimated it in the fifth quintile (calculated probability ≥ 0.32).
Calibration is summarized Figure 2. The slope of the linear predictor (calibration slope) was 0.85 (95% CI: 0.53 to 1.17) indicating that the calculated probabilities were slightly optimistic: low ones are too low and high ones somewhat are too high. The calibration intercept was -0.16 (95% CI: -0.59 to 0.28).

The updated final model is summarized in Table II.

**DISCUSSION**

We developed a prediction model to calculate pregnancy chances during the whole IVF process, both for the first cycle as well as after one or more failed cycles, and taking both fresh embryo transfer and frozen-thawed embryo transfers into account. The model was developed using a careful blockwise building strategy with data systematically collected in consecutive IVF patients. The resulting model produced a range of calculated probabilities, that were well calibrated, both in the development set and in a separate validation set, which contained data that had not been used for model construction.

We used data for the development of the model that were collected during a period of 8 years. Changes in indications for IVF and IVF practice could have affected the influence of predictive factors over time but validation in a more recent patient cohort showed similar discrimination and good calibration, compared with the development set.

Live birth as the main outcome for our model would have been ideal. Unfortunately we did not have these data for all the included cycles. Since only 1% to 2% of all ongoing pregnancies result in late miscarriage or stillbirth, we do not expect that our model would fundamentally change and we therefore feel that ongoing pregnancy rate is the second best outcome (41).

As we used data of a single centre only, the generalizability of the model to other clinics has to be evaluated more extensively in future studies (geographical validation). Centers with higher success rates or those with less positive results could use the model after recalibration, by adjusting the intercept to reflect the IVF success rates in their center (42). Such periodic reassessment may also be beneficial within centers, to ensure that calibration is maintained. We have not yet evaluated its impact in counselling individual couples, which is also a topic for additional research.

As is the case for other fertility prediction models discrimination was less than perfect for our model, expressed by the area under the ROC-curve (0.68), but the calibration data showed that the model distinguishes well between couples with a poor, moderate and good prognosis in successive IVF cycles. We feel these data on calibration are more relevant for decision-making than discrimination statistics in the assessment of any fertility
Individualized decision-making in IVF: calculating the chances of pregnancy.

prediction model. Couples undergoing infertility treatment are not concerned about their chances relative to other couples - which is expressed by discrimination - but worry more about their chances of getting pregnant themselves, which is expressed more adequately by calibration (43;44). The calibration was somewhat less optimal in the last two quintiles in the validation set. Couples in the corresponding subgroups have a good prognosis, and can be clearly distinguished from couples with a moderate or poor prognosis. We therefore think that this suboptimal calibration has no real practical relevance as couples with a good prognosis will continue treatment despite a slightly higher or lower probability.

Since the birth of Louise Brown in 1978, the number of in vitro fertilization cycles has increased rapidly: in the United Kingdom there were 6,650 cycles in 1991 and 57,652 cycles in 2010 (45;46). This increase is not caused by a sudden epidemic of infertility but by increased access to IVF and by expansion of the indications for IVF. At first IVF was only initiated in couples with bilateral tubal occlusion while later on IVF was also initiated in couples with unexplained subfertility, male subfertility, cervical factor, failed ovulation induction, endometriosis, or unilateral tubal pathology (6;47;48). The major difference between the original indication and the indications for which IVF is conducted nowadays, is that the couples with bilateral tubal pathology have a zero chance of natural conception and completely depend on IVF for getting pregnant, whilst couples with the newer indications are subfertile, and do have chances of natural conception, which may or may not be better than with IVF. For them, these chances have to be balanced against those with IVF. As IVF can be stressful physically and emotionally and is not without health risks, subfertile couples should thus be well informed about the chances for success with IVF before each cycle. Unfortunately at this point there are no randomized controlled clinical trials comparing IVF with natural conception. Thus, the only way to counsel couples properly is by model-based prognosis.

In the current financial climate, healthcare systems all over the world face dramatic budgets cuts. In the United States alone these cuts in healthcare cost are expected to amount to 100 billion US dollars, the National Healthcare System in the United Kingdom has to reduce its budget by 20 billion pounds and in the Netherlands these costs reductions are calculated around 5 billion euros (49-51). With these announced cuts, IVF budget will inevitably be hurt as well. Every fertility specialist should be encouraged to control IVF cost by selecting only those couples for IVF that have a reasonable chance of success and outweighs the burden and health risks of the treatment. At this point the only way to select couples for IVF is selection based on model-based prognosis. Our model enables an individualized calculation of the chances of ongoing pregnancy with IVF. Based on a couple’s specific probability, one can decide whether the chances of an ongoing pregnancy with IVF justify the burden, risks and costs of the treatment.
The use of prediction models in deciding whether couples should receive fertility treatment out of public funding is not new. In the Netherlands and New Zealand prediction models are used to decide which couples would truly benefit from fertility treatment, i.e. fertility treatment indeed increases their chance of conception compared with natural conception, and which couples will not benefit (52).

Before implementing our model in clinical practice, the threshold at which probability to start or to continue treatment should be determined, as this may differ between different stakeholders. To achieve optimal implementation of the model, as shown by a previous implementation study, adequate patient information material should be developed, organization of regular fertility meetings is necessary, the development of local protocols need to be further stimulated and the knowledge and communication skills of professionals ought to be improved (53).

We believe that the IVF model presented here is the first to calculate the chances of an ongoing pregnancy with IVF, both for the first cycle and after any number of failed cycles. Incorporating the model in counselling couples considering IVF may strengthen evidence-based, individualized decision-making and a rational use of scarce resources.
Individualized decision-making in IVF: calculating the chances of pregnancy.

REFERENCES

Chapter 4


Individualized decision-making in IVF: calculating the chances of pregnancy.


Supplementary table

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Universal model</th>
<th>Model Cycle 1</th>
<th>Model Cycle 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cycle 1 and 2 included</td>
<td>2150 cycles</td>
<td>1421 cycles</td>
</tr>
<tr>
<td>-2 Log Likelihood*</td>
<td>2369</td>
<td>1596</td>
<td>768</td>
</tr>
<tr>
<td>Intercept</td>
<td>28.72</td>
<td>30.08</td>
<td>13.12</td>
</tr>
</tbody>
</table>

### Patient characteristics

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>95% CI</th>
<th>β</th>
<th>95% CI</th>
<th>β</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-2.83</td>
<td>(-5.01 - 0.65)</td>
<td>-2.90</td>
<td>(-5.55 - 0.25)</td>
<td>-1.41</td>
<td>(-6.06 - 3.24)</td>
</tr>
<tr>
<td>Age²</td>
<td>0.09</td>
<td>(0.03 - 0.16)</td>
<td>0.09</td>
<td>(0.01 - 0.18)</td>
<td>0.05</td>
<td>(-0.09-0.19)</td>
</tr>
<tr>
<td>Age³</td>
<td>0.00</td>
<td>(0.00 - 0.00)</td>
<td>0.00</td>
<td>(0.00 - 0.00)</td>
<td>0.00</td>
<td>(0.00-0.00)</td>
</tr>
<tr>
<td>Duration of subfertility†</td>
<td>-0.11</td>
<td>(-0.19 - -0.03)</td>
<td>-0.14</td>
<td>(-0.23 - -0.05)</td>
<td>0.00</td>
<td>(-0.16-0.15)</td>
</tr>
<tr>
<td>Previous ongoing pregnancy</td>
<td>0.26</td>
<td>(0.02 - 0.49)</td>
<td>0.30</td>
<td>(0.01 - 0.59)</td>
<td>0.38</td>
<td>(-0.08-0.85)</td>
</tr>
<tr>
<td>Male subfertility</td>
<td>1.19</td>
<td>(-0.44 - 2.83)</td>
<td>0.25</td>
<td>(-1.80 - 2.29)</td>
<td>2.21</td>
<td>(-0.99-5.41)</td>
</tr>
<tr>
<td>Diminished ovarian reserve</td>
<td>-1.01</td>
<td>(-1.59 - -0.44)</td>
<td>-1.16</td>
<td>(-2.00 - -0.33)</td>
<td>-0.95</td>
<td>(-2.06-0.17)</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>-0.62</td>
<td>(-1.16 - -0.07)</td>
<td>-0.99</td>
<td>(-1.81 - -0.16)</td>
<td>-0.18</td>
<td>(-1.08-0.71)</td>
</tr>
<tr>
<td>Basal FSH‡</td>
<td>-0.09</td>
<td>(-0.17 - -0.01)</td>
<td>-0.09</td>
<td>(-0.20 - 0.03)</td>
<td>-0.12</td>
<td>(-0.30-0.07)</td>
</tr>
<tr>
<td>Number of previous failed IVF cycles</td>
<td>-0.26</td>
<td>(-0.46 - -0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age * male subfertility</td>
<td>-0.04</td>
<td>(-0.08 - 0.01)</td>
<td>-0.01</td>
<td>(-0.07 - 0.05)</td>
<td>-0.06</td>
<td>(-0.15-0.03)</td>
</tr>
<tr>
<td>Endometriosis * diminished ovarian reserve</td>
<td>1.96</td>
<td>(0.43 - 3.50)</td>
<td>2.74</td>
<td>(0.13 - 5.36)</td>
<td>2.05</td>
<td>(0.26-4.35)</td>
</tr>
</tbody>
</table>

### Embryo parameters

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>95% CI</th>
<th>β</th>
<th>95% CI</th>
<th>β</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo yes/no after ovum pick up</td>
<td>0.93</td>
<td>(0.07 - 1.80)</td>
<td></td>
<td></td>
<td>0.73</td>
<td>(-0.37-1.82)</td>
</tr>
<tr>
<td>Number of embryos after ovum pick up¥</td>
<td>0.07</td>
<td>(0.01 - 0.12)</td>
<td></td>
<td></td>
<td>0.05</td>
<td>(-0.01-0.12)</td>
</tr>
<tr>
<td>Mean morphological score all embryos day 3</td>
<td>-0.40</td>
<td>(-0.73 - -0.06)</td>
<td></td>
<td></td>
<td>-0.30</td>
<td>(-0.70-0.09)</td>
</tr>
<tr>
<td>Eight-cell embryo yes/no on day 3</td>
<td>-0.36</td>
<td>(-0.70 - -0.03)</td>
<td></td>
<td></td>
<td>-0.37</td>
<td>(-0.78-0.04)</td>
</tr>
<tr>
<td>Morulae yes/no on day 3</td>
<td>0.68</td>
<td>(0.15 - 1.21)</td>
<td></td>
<td></td>
<td>0.48</td>
<td>(-0.17-1.13)</td>
</tr>
</tbody>
</table>

*There is no significant difference between the -2 Log likelihood of the universal model and the -2 Log likelihood of Model cycle 1 and 2 together (p 0.9)

Age² = Age squared

Age³ = Age to the power of 3

† duration of subfertility ≥ 5 years = 5 years

‡ bFSH ≤ 10 IE/L = 10 IE/L

¥ number of embryos ≥ 10 = 10 embryos
Prediction models in in vitro fertilization; where are we?

L.L. van Loendersloot
S. Repping
P.M.M. Bossuyt
F. van der Veen
M. van Wely

Journal of Advanced Research 2013; in press; May 9
Chapter 5

ABSTRACT

Since the introduction of in vitro fertilization (IVF) in 1978, over 5 million babies have been born worldwide using IVF. Contrary to the perception of many, IVF does not guarantee success. Almost 50% of couples that start IVF will remain childless, even if they undergo multiple IVF cycles. The decision to start or pursue with IVF is challenging due to the high cost, the burden of the treatment, and the uncertain outcome. In optimal counselling on chances of a pregnancy with IVF, prediction models may play a role, since doctors are not able to correctly predict pregnancy chances. There are three phases of prediction model development: model derivation, model validation, and impact analysis. This review provides an overview on predictive factors in IVF, the available prediction models in IVF and provides key principles that can be used to critically appraise the literature on prediction models in IVF. We will address these points by the three phases of model development.
Prediction models in in vitro fertilization; where are we?

INTRODUCTION

Since the birth of Louise Brown in 1978, over 5 million babies have been born worldwide using in vitro fertilization (IVF) (1). The number of in vitro fertilization cycles has increased rapidly; in 2006, 458,759 cycles were reported in 32 European countries, 99,199 cycles in the USA and 50,275 cycles in Australia and New Zealand (2-4). The number of cycles is increasing each year even further.

The increase in IVF cycles is not caused by a sudden epidemic of infertility, but by increased access to IVF, and by an expansion of the indications for IVF. Initially, IVF was performed in couples with bilateral tubal occlusion (5). In 1992, intracytoplasmic sperm injection (ICSI) was first introduced and initiated in couples with severe male subfertility (6). Later on, IVF/ICSI was also applied in couples without an absolute indication for IVF, such as unexplained subfertility, cervical hostility, failed ovulation induction, endometriosis, or unilateral tubal pathology (7;8). The major difference between the original indication and the indications for which IVF is conducted nowadays is that the couples with bilateral tubal pathology or severe male subfertility have a zero chance of natural conception and completely depend on IVF/ICSI for a pregnancy, while couples with the newer indications are subfertile: they do have chances of natural conception, which may or may not be better than with IVF.

Despite the lack of evidence that IVF is effective in couples without an absolute IVF indication, IVF is often considered as a last resort for all subfertile couples regardless of the etiology of their subfertility (7-12). Contrary to the perception of many, IVF does not guarantee success; almost 38-49% of couples that start IVF will remain childless, even if they undergo six IVF cycles (13). Subfertile couples should therefore be well informed about the chances of success with IVF before starting their first or before continuing with a new IVF cycle. Based on a couple’s specific probability, one should decide whether the chances of success with IVF justify the burden, risks, and costs of the treatment. The threshold at which probability to start or to continue treatment may differ between different stakeholders, such as insurance companies, the tax payer, and the patients.

In optimal counselling on chances of a pregnancy after IVF, pregnancy prediction models may play a role, since doctors are not able to correctly predict pregnancy chances(14;15). Predictions made by clinicians on the basis of clinical experience or “gut-feeling” have only slight to fair reproducibility, indicating that these predictions are likely to be inaccurate (15).

The efforts to develop prediction models for IVF reflect the need for such models in clinical practice. This need can be explained by the inability of diagnostic tests to detect factors that indicate subfertility with near 100% certainty in patients. Accurate diagnostic tests would allow treatment to focus on specific factors (16). As IVF is currently used as an
empirical treatment and not as a causal intervention for a specific disorder, there is a strong need to distinguish between couples with a good and a poor prognosis (16). In the absence of randomized clinical trials, evaluating the effectiveness of IVF prediction models can be used to counsel couples.

The development of a prediction model can be divided into three phases: model derivation, model validation, and impact analysis (Fig. 1) (16;17). In the model derivation phase, predictors are identified, based on prior knowledge, and the weight of each predictor (regression coefficient) is calculated. In the model validation phase, the performance of the model, i.e. model’s ability to predict outcome is evaluated, and also the “generalizability” or “transportability” of the model is evaluated. The third and final phase consists of impact analysis. The impact analysis establishes whether the prediction model improves doctors’ decisions by evaluating the effect on patient outcome (16;17).

Figure 1 | Three phases of model development

This review provides an overview on predictive factors in IVF, the available prediction models in IVF and provides key principles that can be used to critically appraise the literature on prediction models in IVF. We will address these points by the three phases of model development: model derivation, model validation, and impact analysis.

Phase 1: model derivation

Identification of predictors

Candidate predictors are variables that are chosen to be studied for their predictive performance. These can include subject demographics, clinical history, physical examination, disease characteristics, test results, and previous treatments (18). The identification of candidate predictors is preferably based on subject knowledge, on pathophysiological mechanisms, or the results of previous studies. Studied predictors should be clearly defined, standardized, and reproducible to enhance generalizability and application of study results to practice (18). Researchers frequently measure more predictors than can reasonably be analysed. When the number of predictors is much larger than the number of outcome events, there is a risk of overestimating the predictive performance of the model. To reduce the risk of false positive findings (predictors), at least 10 individuals having (developed) the event of interest are needed per candidate variable/predictor to allow for reliable prediction modelling (19).
Prediction models in in vitro fertilization; where are we?

A recent systematic review and meta-analysis on predictive factors in IVF evaluated nine predictive factors: female age, duration of subfertility, type of subfertility, indication for IVF, basal follicle stimulating hormone (bFSH), fertilization method, number of oocytes, number of embryos transferred, and embryo quality (20).

Female age is one of the most important prediction factors for success with IVF. Increasing female age was associated with lower pregnancy chances in IVF (OR 0.95, 95% CI: 0.94-0.96) (20). The decrease in fertility sets in after the age of 30 years, with a marked decline after 35 years for both spontaneous as IVF-induced pregnancies (20-23). The biological explanation for the declining chances to conceive with increasing female age most likely lies in the diminished ovarian reserve: the decrease in both quantity and quality of oocytes (24). Diminished ovarian reserve generally leads to a poor response to gonadotropin therapy and limits the possibility of a successful pregnancy (25).

Increasing duration of subfertility is known to be associated with a reduced possibility of natural conception (adjusted hazard rate 0.83; 95% CI 0.78-0.88) (7;26-30). In IVF, pregnancy rates were slightly lower in couples with a longer duration of subfertility (OR 0.99, 95% CI: 0.98-1.00), even after adjustment for age (20;23;31-33).

Although the meta-analysis did not find a significant association between type of subfertility (primary versus secondary subfertility) and pregnancy with IVF (unadjusted OR 1.04 95% CI: 0.65-1.43), two recent, large studies did find an association (20;31;33). A previous ongoing pregnancy or live birth, adjusted for factors such as age, substantially increases the likelihood of success with IVF (31;33).

Through the years, several studies have reported on the association between the indication for IVF and pregnancy with IVF without consistent results. These studies did not use the same reference categories making the interpretation of the data difficult. There is evidence for an association between tubal pathology and pregnancy with IVF. Women with tubal pathology alone had lower pregnancy chances compared to women with unexplained subfertility or other indications (23;31;34-36). On the other hand, another study suggested that women with tubal pathology had higher pregnancy chances after IVF compared with couples with unexplained subfertility, though not significantly (37). There is also evidence for an association between male subfertility and pregnancy with IVF. Although two studies (N=2,628 cycles) reported that couples with male subfertility have lower pregnancy chances than those with unexplained subfertility a very large cohort study (N=144,018 cycles) showed that couples with only male subfertility had increased pregnancy chances compared to couples with unexplained subfertility (31;35;36). Since these studies use different reference categories and different number of categories, it is not possible to compare these results optimally. For future studies and the development for prediction models, it would be useful to report every indication for IVF as a separate variable instead of combining all indications into one factor, to be able to compare all studies (20).
Basal FSH is an indirect estimate of ovarian reserve. A higher bFSH value was associated with lower pregnancy rates after IVF (OR 0.94; 95% CI: 0.88-1.00) (20).

Increasing number of oocytes was associated with higher pregnancy chances with IVF (OR 1.04, 95% CI: 1.02-1.07) (20). A recent large cohort study (N= 400,135) also showed a strong relationship between the number of oocytes and live birth rate with IVF. The association is not linear; the best chance of live birth is associated with approximately 15 oocytes (38).

Although the meta-analysis did not find a significant association between pregnancy chances with ICSI compared to IVF (OR 0.95, 95% CI: 0.79-1.14), a more recent large cohort study (N=144,018 cycles) reported higher chances with ICSI compared to IVF (OR 1.28, 95% CI: 1.25-1.31), even after adjusting for all relevant factors (OR 1.27, 95% CI: 1.23-1.31) (20;31).

The number of embryos transferred and embryo quality were associated with increased pregnancy chances (20).

Estimation of the regression coefficient
After identifying all potential predictors, a multivariable model can be constructed by regression analysis (logistic regression or proportional hazard analysis). To evaluate the quantitative effect of each predictor, the weight of each predictor is calculated by estimating the corresponding regression coefficient in a linear model.

Currently, over 21 papers have reported on the development and or validation of models for the prediction of pregnancy with IVF (Table I) (23;31-37;39-55).

Phase 2: model validation
The second phase in the development of a prediction model is the evaluation of the model performance, i.e. model validation. The performance of the model can be evaluated by calculating its discriminative capacity and the degree of calibration. Discrimination relates to how well a model can distinguish between patients with and without the outcome, i.e. discriminate between women who achieved pregnancy and those who did not. Discriminative capacity can be expressed by the area under the receiver operating characteristic curve (AUC), also known as the c-statistic. A model with a c-statistic of 0.5 has no discriminative power at all, while 1.0 would reflect perfect discrimination. Calibration relates to the agreement between observed outcomes and calculated probabilities, i.e. if we calculate a 30% probability of a pregnancy with IVF, the observed relative frequency of pregnancy should be approximately 30 out of 100 women. Calibration can be assessed by the Hosmer and Lemeshow goodness-of-fit test statistic. A Hosmer–Lemeshow statistics with a p-value above 0.05 implies that there is no significant miscalibration. In addition, calibration can also be assessed by comparing
Prediction models in in vitro fertilization; where are we?

The average calculated probabilities with the actual proportions in disjoint subgroups. The average calculated probabilities and actual proportions in each group can be plotted in a calibration plot. In case of perfect calibration, all points in a calibration plot are on the diagonal, the line of equality, and probabilities correspond perfectly to the actual proportions.

Table I | Characteristics of prediction models for pregnancy after IVF and IVF-eSET

<table>
<thead>
<tr>
<th>First author (year)</th>
<th>Inclusion of embryo characteristics</th>
<th>IVF-eSET</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Loendersloot et al. (2013)</td>
<td>Yes</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Nelson et al. (2011)</td>
<td>No</td>
<td>No</td>
<td>Live birth</td>
</tr>
<tr>
<td>van Weert et al. (2008)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Lintsen et al. (2007)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Verberg et al. (2008)</td>
<td>Yes</td>
<td>Yes</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Carrera-Rotlan et al. (2007)</td>
<td>No</td>
<td>No</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>Ottoson et al. (2007)</td>
<td>Yes</td>
<td>Yes</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>Ferlitsch et al. (2004)</td>
<td>No</td>
<td>No</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>Hunault et al. 2002</td>
<td>Yes</td>
<td>Yes</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Stolwijk et al. (2000)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Bancsi et al. (2000)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Minaretzis et al. (1998)</td>
<td>Yes</td>
<td>No</td>
<td>Live birth</td>
</tr>
<tr>
<td>Commenges-Ducos et al. (1998)</td>
<td>Globel model: No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Model for implantation: Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Templeton et al. (1996)</td>
<td>No</td>
<td>No</td>
<td>Live birth</td>
</tr>
<tr>
<td>Stolwijk et al. (1996)</td>
<td>Model A: No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td></td>
<td>Model B: Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model C: Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bouckaert et al. (1994)</td>
<td>Yes</td>
<td>No</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>Haan et al. (1991)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Hughes et al. (1989)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
<tr>
<td>Nayuda et al. (1989)</td>
<td>No</td>
<td>No</td>
<td>Ongoing pregnancy</td>
</tr>
</tbody>
</table>

The validation phase can be subdivided in internal validation (phase 2a) and external validation (phase 2b). With internal validation, the model’s ability to predict the outcome in the group of patients in which it was developed is evaluated (reproducibility). Internal validation should be seen as validating the modeling process (56). Of the 21 papers reporting on IVF prediction model development, only 11 are also internally validated (23;31-35;37;40;45;49-51;53-55).
Before being able to use prediction models for clinical decision making it is not enough to demonstrate a reasonable or good performance after internal validation. Most models show too optimistic results, even after corrections from interval validation procedures. It is essential to confirm that any developed model also predicts well in a “similar but different” population outside the development set, i.e. external validation (generalizability). The more these populations differ from the development study, the stronger the test of generalizability of the model (57).

There are three different types of external validation, temporal validation, geographical validation, and domain validation. In temporal validation, the model is validated on new patients that are from the same center as the development set, but in a different time period (57;58). In geographical external validation, the model is validated on new patients from a different center as the development set (57;58). In domain validation, the model is validated on new patients that are very different from the patients from which the model was developed (57).

Of the 11 IVF models that went through internal validation, only four models have also been validated externally (32;33;37;45;49-51;53). One model was validated temporally, the model calibrated well both in the development set and in a separate validation set (33). Three models have been validated geographically, but only one model showed good calibration after validation (32;37;45;49-51;53). So at this moment, there is only one model that is generalizable to other clinics (37;45). All other models have to be geographically validated first before using the models in practice.

A prediction model often performs less well in a new group of patients than in the study group in which it was developed. This can be caused by differences in the case-mix between the development and validation population or by true differences between populations (58). Instead of simply rejecting the prediction model and develop or fit a new one, a better alternative is to update existing prediction models and adjust or recalibrate it to the local circumstances or setting of the validation set (57;58). As a result, the updated model is adjusted to the characteristics of new individuals. Several methods for updating prediction models are possible. Most often, differences are seen in the outcome frequency between the development and new validation set. This results in poor calibration of the model; predicted probabilities are systematically too high or too low. By adjusting the intercept (baseline risk) of the original model, calibration can be improved. Additional updating methods vary from adjustment of all predictor regression coefficients, adjustment of regression coefficients for particular predictor weight, to the addition of a completely new predictor or marker to the existing model (57;58).

As patient populations may shift during the years, the group of patients used for the development and validation of the prediction model may differ from the current patient population. Reproductive techniques may evolve during the years, new biomarkers with
Prediction models in in vitro fertilization; where are we?

predictive value may become available, and the prediction model should be regularly updated and adapted to the new setting, so that predictions for future patients remain valid and may even improve (58). IVF centers should therefore consider collecting their own data in electronic databases, so that with accumulation of the number of IVF cycles over time they can update the model with their own data.

Phase 3: impact analysis
The third and final phase in the evaluation of models is impact analysis; it establishes whether the prediction model improves decisions, in terms of quality or cost-effectiveness of patient care (17;57;58). This can be evaluated in one setting (phase 3a) or in varied settings (phase 3b). Different study designs to evaluate the impact of a prediction model are possible, such as comparing the outcomes between patients randomly assigned to receive management guided by the prediction model and patients managed without the prediction model (care-as-usual). A less valid alternative is to ask fertility specialists to document therapeutic management decisions before and after being “exposed” to a model’s predictions. None of the existing IVF prediction models has reached the impact analysis phase yet.

DISCUSSION
As IVF can be stressful physically and emotionally and is not without health risks, subfertile couples should thus be well informed about the chances for success with IVF before each cycle. Unfortunately at this point, there are no randomized controlled clinical trials comparing IVF with natural conception. Thus, the only way to counsel couples properly is by model-based prognosis.

Over 21 papers have reported on the development and/or validation of prediction models in IVF. Of these 21 papers, only two models had a good performance after external validation. Impact analyses have not yet been performed for any of these models. Future research should focus more on updating existing prediction models and adjust or recalibrate them to the local circumstances or setting rather than developing new prediction models. This way prediction models may strengthen evidence-based, individualized decision-making and can contribute to a rational use of scarce resources.
REFERENCES


Prediction models in in vitro fertilization; where are we?


Prediction models in in vitro fertilization; where are we?


PART TWO
Optimizing embryo transfer strategies
Pregnancy and twinning rates using a tailored embryo transfer policy

L.L. van Loendersloot
M. van Wely
M. Goddijn
S. Repping
P.M.M. Bossuyt
F. van der Veen

Chapter 6

ABSTRACT

This study prospectively evaluates a tailored embryo transfer policy based on the prognostic profile of the couple. Single embryo transfer (SET) was performed followed by double embryo transfer (DET) in frozen embryo transfer cycles in women with a good prognosis (aged <35 years, first cycle, ≥1 top quality embryo). DET was performed in both fresh and frozen cycles in women with an intermediate prognosis (<35 years, first cycle and no top quality embryo available, or aged <35 years and ≥1 failed cycles, or aged 35-38 years). Triple embryo transfer (TET) in both fresh and frozen cycles was performed in women with a poor prognosis (aged ≥39 years). The cumulative ongoing pregnancy rate in women with a good prognosis was 43%, with a multiple pregnancy rate of 2%. In women with an intermediate prognosis this was 27% and 23%, respectively. Corresponding rates were 18% and 13% in women with a poor prognosis. The data in this study can be used to guide current practice, i.e. performing SET in women with a good prognosis and TET in women with a poor prognosis. The embryo transfer strategy in women with an intermediate prognosis requires further improvement, possibly by refining the prognosis according to the ovarian response after ovarian stimulation.
Pregnancy and twinning rates using a tailored embryo transfer policy.

INTRODUCTION

In the last decades improvements in in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI), have resulted in higher embryo implantation rates (1). These higher implantation rates in combination with the transfer of multiple embryos led to a substantial increase in multiple pregnancy rates (2-4). These high multiple pregnancy rates caused great concern, since maternal and perinatal morbidity and mortality as well as health care costs became unacceptable (1;5;6). As the number of embryos transferred is the most important factor influencing multiple pregnancy rates, the only way to reduce these high multiple pregnancy rates was to reduce the number of transferred embryos.

The first randomized trial to compare elective single embryo transfer (eSET) and double embryo transfer (DET) was performed in 1999 in women under 34 undergoing their first IVF/ICSI cycle (7). Since then, two systematic reviews and meta-analyses of randomized trials have synthesized aggregated data and individual patient data on eSET versus double embryo transfer (DET) in women with a good prognosis. These reviews showed that eSET not only reduces the odds of multiple pregnancies but almost halves the odds of a live birth per fresh cycle. These reduced live birth rates after eSET were restored by the subsequent transfer of a single frozen thawed embryo yielding cumulative live birth rates comparable to those after DET (8;9).

The implementation of single embryo transfer policy in countries such as Sweden and Belgium for women with a good prognosis resulted in a dramatic decrease in multiple pregnancy rates. In Sweden there was a decrease from 35% to around 5% and in Belgium from 19% to 3%, while maintaining similar pregnancy rates (10;11).

Although eSET is now an accepted policy for women with a good prognosis, the majority of women currently undergoing IVF have an intermediate or poor prognosis, such as women over 35 years with one or more failed IVF cycles. For these women data from randomized trials are lacking and existing cohort studies have not been able to provide robust evidence on how many embryos to transfer to obtain high pregnancy rates at low multiple pregnancy rates. This is because only women with good response after ovarian stimulation or women with at least two or more good quality embryos have been included in the cohort studies conducted thusfar (12;13).

In view of this lack of data on this important issue we here report on the results of a prospective cohort study on the implementation of a differentiated embryo transfer policy based on the age of the woman, the number of previous cycles and embryo quality, in terms of ongoing pregnancy rates and multiple pregnancy rates.
MATERIAL AND METHODS

In 2006 we implemented a tailored embryo transfer strategy as our standard clinical care and prospectively monitored the effects of this strategy on relevant clinical outcomes. We evaluated all consecutively performed IVF/ICSI cycles in the Academic Medical Center between August 2006 and April 2011.

All couples had been trying to conceive for at least 12 months and underwent a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (14). The indication to start IVF or ICSI treatment was determined according to the Dutch IVF guideline (15). If subfertility was caused by tubal pathology, such as two-sided tubal blockage and severe endometriosis, or severe oligozoospermia (post-wash total motile sperm count < 3 million) IVF/ICSI was offered directly (16). In case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia, and unexplained subfertility, at least six intra uterine inseminations (IUI) were applied before IVF/ICSI was offered. In case of ovulation disorders, mainly caused by polycystic ovary syndrome (PCOS), 12 cycles of ovulation induction were applied before IVF/ICSI was offered.

The embryo transfer policy was based on the prognostic profile of the women. The prognostic profiles, i.e. good-/intermediate-/poor prognosis, were based on three important predictive factors for pregnancy with IVF: female age, number of previous cycles and embryo quality (17-19).

Women with a good prognosis were women under 35 years undergoing their first cycle of IVF/ICSI with at least one top-quality embryo. In these women a single embryo transfer (SET) was performed. In the frozen embryo transfer cycles (FET) following these fresh cycles, double embryo transfer (DET) was performed (Table I).

Women with an intermediate prognosis were women under the age of 35 who did not have a top quality embryo in the first cycle, or women under the age of 35 who failed to get pregnant in their first cycle of IVF/ICSI, or women between 35 and 38 years of age. In these women DET was performed in the fresh and frozen cycles (Table I). This transfer strategy was based on a combination of the Practice Committee of the American Society for Reproductive Medicine guidelines and the Belgian embryo transfer legislation (20;21).

Women with a poor prognosis were women of 39 years or older. In these women three embryos were transferred (TET) in the fresh and frozen cycles (Table I). This strategy was also based on a combination of the Practice Committee of the American Society for Reproductive Medicine guidelines and the Belgian embryo transfer legislation (20;21).
Pregnancy and twinning rates using a tailored embryo transfer policy.

Table 1 | Embryo transfer strategy

<table>
<thead>
<tr>
<th>Good prognosis</th>
<th>Fresh embryo transfer</th>
<th>Frozen and thawed embryo transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women &lt; 35 years and first IVF/ICSI cycle</td>
<td>SET*</td>
<td>DET†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intermediate prognosis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Women &lt; 35 years first IVF cycle and no top quality embryo available</td>
<td>DET†</td>
<td>DET†</td>
</tr>
<tr>
<td>Women &lt; 35 years and ≥ 1 failed IVF cycle(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women 35-38 years</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Poor prognosis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Women ≥39 years</td>
<td>TET‡</td>
<td>TET‡</td>
</tr>
</tbody>
</table>

*SET: transfer of a single top-quality embryo
†DET: double embryo transfer
‡TET: triple embryo transfer

Fresh and frozen-thawed IVF/ICSI cycles were included for analysis if the embryo transfer criteria were met: women eligible for SET had at least one top quality embryo, women eligible for DET had at least two embryos and women eligible for TET had at least three embryos.

**IVF/ICSI procedures**

Women underwent controlled ovarian hyperstimulation after down-regulation with the GnRH agonist triptorelin (Decapeptyl®, Ferring) in a long protocol with a midluteal start. Controlled ovarian hyperstimulation was started on cycle day 5 with recombinant FSH or HMG in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Follicular maturation was induced by 10,000 IU human chorionic gonadotropin hormone (hCG) (Pregnyl, Organon). Cumulus-oocyte complexes were recovered by transvaginal ultrasound guided follicle aspiration 36 hours thereafter. Oocytes were inseminated with 10,000 or 15,000 progressively motile spermatozoa (in vitro fertilization) or injected with a single spermatozoon (intracytoplasmic sperm injection) 2-4 hours after follicle aspiration. Embryos were cultured in Human Tubal Fluid (HTF, Gynotec, Malden, The Netherlands or G5 medium, Vitrolife, Sweden) at 37°C and 5% CO₂ in air. Embryo transfer was performed on day 3 after follicle aspiration with a Wallace catheter (Smiths Medical, UK). Supernumerary embryos of good quality were frozen on day 4 after follicle aspiration using a slow-freeze protocol. The luteal phase was supported by 200 mg progesterone intravaginally (Utrogestan, Besins International) twice per day. An hCG blood test was performed 18 days after oocyte retrieval.
Morphological scoring
All embryos were cultured individually. On day 2 and day 3 the number of blastomeres was assessed and each embryo was given a morphological score. For the morphological score the degree of fragmentation of the embryo and the uniformity of the blastomeres were assessed (22). The embryo was given a score of 1 (>50% fragmentation), 2 (20-50% fragmentation), 3 (<20% fragmentation) or 4 (no fragmentation). If the blastomeres of the embryo were non-uniform in size the morphological score was augmented with one point with 1 remaining the lowest possible score. If on day 3 the embryo showed signs of compaction the embryo was scored as a morula and given a grade based on the degree of compaction (score 1: full compaction, score 2: 50-<100% compaction and score 3: less then 50% compaction). A cumulative embryo score was calculated by multiplying the number of blastomeres with the corresponding morphological score (23). For example, an 8-cell embryo with a morphological score of 3, would receive a cumulative score of 24 (8 x 3). An 4-cell embryo with a morphological score of 4, would receive a cumulative score of 16 (4 x 4). Morulae were considered top-quality if at least 50% of the cells were part of the compacting process. Top quality embryos were defined as embryos with a cumulative embryo score of 24 or higher on day 3 after follicle aspiration.

Outcomes
For each transfer we observed whether or not the transfer resulted in an ongoing pregnancy and, if so, in a single or a multiple pregnancy. Ongoing pregnancy was defined as a viable intra-uterine pregnancy of at least 11 weeks.

Statistical analysis
All IVF/ICSI cycles were eligible for our evaluation, but whenever less or more embryos were transferred than indicated by the protocol, the corresponding cycles were excluded from the analyses (per protocol analysis). We calculated ongoing pregnancy rates and multiple pregnancy rates per cycle in each of the three prognostic categories. Ongoing pregnancy rates were defined as the number of ongoing pregnancies relative to the number of transfers. Multiple pregnancy rates were defined as the number of multiple ongoing pregnancies relative to the number of ongoing pregnancies. Cumulative ongoing pregnancy rates were defined as the number of pregnancies obtained by fresh embryo transfer or subsequent embryo transfer(s) of frozen-thawed embryos from the same IVF/ICSI cycle relative to the number of fresh transfers. If in the same cycle women got pregnant both after fresh and after frozen-thawed embryo transfer(s), this was reported as a single pregnancy in the calculation of the cumulative continuing pregnancy rate. Chi-squared test and ANOVA statistics were used to evaluate differences between the baseline characteristics and the outcome measures in all categories. If there were less than five (multiple) pregnancies Fisher's Exact test was used. Differences were classified as statistically significant at P < 0.05. Statistical analysis was performed using the Statistical Package for Social Sciences (SPSS) version 18.0 software (SPSS, Chicago, IL, USA). We reported all statistics in a per cycle analysis.
Pregnancy and twinning rates using a tailored embryo transfer policy.

Under the legal requirements for clinical research in the Netherlands, this study was exempt from medical ethics and institutional review board (IRB) approval.

RESULTS

In total 3,606 cycles of IVF/ICSI were performed in 1896 women. Three hundred and seventy-five cycles (10.4%) were cancelled prior to follicle aspiration. In 31 cycles there were no oocytes after follicle aspiration. In 257 cycles there were no embryos available for transfer due to total fertilization failure. In 24 cycles there were no suitable embryos for transfer and in 28 cycles the embryo transfer was cancelled due to medical reasons as ovarian hyperstimulation syndrome or pelvic inflammatory disease. In 302 cycles there were fewer embryos available than required for the embryo transfer protocol. In 260 cycles less embryos and in 113 cycles more embryos were transferred than required by the protocol, because of women’s request, medical reasons or participation in a trial (Fig 1). Two thousand two hundred and sixteen cycles in 1402 women thus remained for analysis. From these fresh IVF/ICSI cycles, 993 frozen-thawed embryo transfers were initiated. In 85 cycles no embryos were available for transfer after thawing. In 309 cycles fewer embryos were available than required by the embryo transfer protocol. In 7 cycles more embryos were transferred than required by the protocol. Five hundred and ninety two embryo transfers were thus available for analysis (Fig 1). On average 1.6 cycles (IQR 1-2) per patient were performed. Given this small number of cycles per patient, we decided not to correct for it in the analysis and report statistics in a per cycle analysis (not in a per woman analysis). The patient characteristics of the included cycles are shown in Table II.

Fresh embryo transfers

Ongoing pregnancy and multiple pregnancy rates are summarized for each prognostic category in Table IIIa. In the cycles of women with a good prognosis the ongoing pregnancy rate was 34% (138/408) per transfer with a multiple pregnancy rate of 1% (1/138). In the cycles of women with an intermediate prognosis the ongoing pregnancy rate was 25% (325/1317) and the multiple pregnancy rate was 25% (80/325). The pregnancy and multiple pregnancy rates for the subcategory of women with intermediate prognosis were as follows: in the cycles of women younger than 35 years with no top quality embryos available 25% (52/206) and 15% (8/52), in the cycles of women younger than 35 years after one or more failed cycles 27% (87/325) and 34% (30/87), and in the cycles of women between 35 and 38 years 24% (186/786) and 23% (42/186). In cycles of women with a poor prognosis the ongoing pregnancy rate was 17% (82/491) and the multiple pregnancy rate was 13% (11/82).
Chapter 6

Figure 1 | Flowchart

Fresh IVF/ICSI cycles 3606

- Number of cycles cancelled prior to oocyte retrieval 375
- No oocytes after follicle aspiration 31
- Number of total fertilisation failures 257
- No suitable embryos for transfer 24
- Cancel embryo transfer due to medical reasons 28

Less embryos available than transfer protocol 302

Less embryos transferred than transfer protocol
- Intermediate prognosis SET instead of DET 128
- Poor prognosis SET instead of TET 6
- Poor prognosis DET instead of TET 126

More embryos transferred than transfer protocol
- Good prognosis DET instead of SET 87
- Intermediate prognosis TET instead of DET 26

Fresh IVF/ICSI cycles 2216
Frozen-thawed embryo cycles 993

- No embryos available for transfer after thawing 85
- Less embryos available than transfer protocol 309
- More embryos transferred than transfer protocol 7
Pregnancy and twinning rates using a tailored embryo transfer policy.

### Table II | The patient characteristics of the included IVF/ICSI cycles

<table>
<thead>
<tr>
<th>Included cycles</th>
<th>Good prognosis</th>
<th>Intermediate prognosis</th>
<th>Poor prognosis</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td><strong>%</strong></td>
<td><strong>N</strong></td>
<td><strong>%</strong></td>
<td><strong>N</strong></td>
</tr>
<tr>
<td>Number of patients</td>
<td>1402</td>
<td>297</td>
<td>822</td>
<td>283</td>
</tr>
<tr>
<td>Number of IVF cycles</td>
<td>2216</td>
<td>408</td>
<td>1317</td>
<td>491</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>35.4 (4.6)</td>
<td>30.7 (3.2)</td>
<td>34.8 (3.5)</td>
<td>41.1 (1.2)</td>
</tr>
<tr>
<td>Indication for IVF/ICSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal pathology</td>
<td>338</td>
<td>15%</td>
<td>61</td>
<td>15%</td>
</tr>
<tr>
<td>Unexplained subfertility</td>
<td>634</td>
<td>29%</td>
<td>67</td>
<td>16%</td>
</tr>
<tr>
<td>Male subfertility</td>
<td>1086</td>
<td>49%</td>
<td>261</td>
<td>64%</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>81</td>
<td>4%</td>
<td>10</td>
<td>2%</td>
</tr>
<tr>
<td>Others</td>
<td>206</td>
<td>9%</td>
<td>42</td>
<td>10%</td>
</tr>
<tr>
<td>Fertilization method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVF</td>
<td>1002</td>
<td>45%</td>
<td>165</td>
<td>40%</td>
</tr>
<tr>
<td>ICSI</td>
<td>1214</td>
<td>55%</td>
<td>243</td>
<td>60%</td>
</tr>
<tr>
<td>Mean number of oocytes (SD)</td>
<td>10.8 (5.3)</td>
<td>11.8 (5.5)</td>
<td>10.7 (5.4)</td>
<td>10.5 (4.9)</td>
</tr>
<tr>
<td>Mean number of embryos (SD)</td>
<td>6.5 (3.7)</td>
<td>7.2 (4.0)</td>
<td>6.3 (3.6)</td>
<td>6.7 (3.6)</td>
</tr>
</tbody>
</table>

NS= not statistically significant

**Cumulative pregnancy rates**

The pregnancy data for fresh and frozen-thawed embryo transfers (Table IIIa, b) were combined in order to calculate the cumulative pregnancy rates (Table IIIc). In the cycles of women with a good prognosis the cumulative ongoing pregnancy rate was 43% (177/408) and the cumulative multiple pregnancy rate was 2% (3/177). In the cycles of women with an intermediate prognosis the cumulative ongoing pregnancy rate was 27% (350/1317) and the cumulative multiple pregnancy rate was 23% (82/350). The pregnancy and multiple pregnancy rates for the subcategories of women with intermediate prognosis were as follows: in the cycles of women younger than 35 years with no top quality embryos available 26% (53/206) and 15% (8/53), in the cycles of women younger than 35 years after one or more failed cycles 28% (92/325) and 34% (31/92), and in the cycles of women between 35 and 38 years 26% (205/786) and 21% (43/205).

In the cycles of women with a poor prognosis the cumulative ongoing pregnancy rate was 18% (88/491) and a multiple pregnancy rate was 13% (11/88).
### Table III | Number of fresh and frozen-thawed embryo transfers, pregnancy rates and multiple pregnancy rates per embryo transfer

#### IIIa. Fresh embryo transfers

<table>
<thead>
<tr>
<th>Female age and cycle number</th>
<th>Number of embryos transferred</th>
<th>Number of cycles</th>
<th>Number of pregnancies</th>
<th>Pregnancy rate</th>
<th>Implantation rate</th>
<th>Number of twin pregnancies</th>
<th>Number of triplet pregnancies</th>
<th>Multiple pregnancy rates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good prognosis</strong>&lt;br&gt; &lt; 35 years, cycle 1</td>
<td>1</td>
<td>408</td>
<td>138</td>
<td>34%</td>
<td>34%</td>
<td>1</td>
<td>0</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Intermediate prognosis</strong>&lt;br&gt; &lt; 35 years, cycle ≥2</td>
<td>2</td>
<td>325</td>
<td>87</td>
<td>27%</td>
<td>18%</td>
<td>30</td>
<td>0</td>
<td>34%</td>
</tr>
<tr>
<td>35-38 years</td>
<td>786</td>
<td>186</td>
<td>24%</td>
<td>15%</td>
<td>41</td>
<td>1</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td><strong>Total intermediate prognosis</strong>&lt;br&gt;</td>
<td></td>
<td>1317</td>
<td>325</td>
<td>25%</td>
<td>15%</td>
<td>79</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Poor prognosis</strong>&lt;br&gt; ≥39 years</td>
<td>3</td>
<td>491</td>
<td>82</td>
<td>17%</td>
<td>6%</td>
<td>11</td>
<td>0</td>
<td>13%</td>
</tr>
</tbody>
</table>

#### IIIb. Frozen-thawed embryo transfers

<table>
<thead>
<tr>
<th>Female age and cycle number</th>
<th>Number of FET</th>
<th>Number of cycles</th>
<th>Number of pregnancies</th>
<th>Pregnancy rate</th>
<th>Implantation rate</th>
<th>Number of twin pregnancies</th>
<th>Number of triplet pregnancies</th>
<th>Multiple pregnancy rates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good prognosis</strong>&lt;br&gt; &lt; 35 years, cycle 1</td>
<td>2</td>
<td>203</td>
<td>39</td>
<td>19%</td>
<td>10%</td>
<td>2</td>
<td>0</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Intermediate prognosis</strong>&lt;br&gt; &lt; 35 years, cycle ≥2</td>
<td>2</td>
<td>84</td>
<td>5</td>
<td>6%</td>
<td>4%</td>
<td>1</td>
<td>0</td>
<td>20%</td>
</tr>
<tr>
<td>35-38 years</td>
<td>196</td>
<td>19</td>
<td>10%</td>
<td>5%</td>
<td>1</td>
<td>0</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td><strong>Total intermediate prognosis</strong>&lt;br&gt;</td>
<td></td>
<td>302</td>
<td>25</td>
<td>8%</td>
<td>4%</td>
<td>2</td>
<td>0</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Poor prognosis</strong>&lt;br&gt; ≥39 years</td>
<td>3</td>
<td>87</td>
<td>6</td>
<td>7%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>
Continuation of table III

<table>
<thead>
<tr>
<th>Female age and cycle number</th>
<th>Number of cycles</th>
<th>Number of pregnancies</th>
<th>Pregnancy rate</th>
<th>Number of multiple pregnancies</th>
<th>Multiple pregnancy rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good prognosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 35 years, cycle 1</td>
<td>408</td>
<td>177</td>
<td>43%</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>Intermediate prognosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 35 years, cycle ≥2</td>
<td>206</td>
<td>53</td>
<td>26%</td>
<td>8</td>
<td>15%</td>
</tr>
<tr>
<td>35-38 years</td>
<td>325</td>
<td>92</td>
<td>28%</td>
<td>31</td>
<td>34%</td>
</tr>
<tr>
<td>Total intermediate prognosis</td>
<td>786</td>
<td>205</td>
<td>26%</td>
<td>43</td>
<td>21%</td>
</tr>
<tr>
<td>Poor prognosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥39 years</td>
<td>491</td>
<td>88</td>
<td>18%</td>
<td>11</td>
<td>13%</td>
</tr>
</tbody>
</table>
Protocol violations

In 373 cycles the embryo transfer policy deviated from the tailored protocol. In the cycles of women with a good prognosis where DET instead of the SET strategy that was dictated by the protocol was performed, the ongoing pregnancy rate was 38% (33/87), not significantly different compared to the ongoing pregnancy rate achieved after SET, and a multiple pregnancy rate of 39% (13/33) which was significantly higher than with SET ($p<0.001$). In the cycles of women with an intermediate prognosis where SET instead of the DET strategy that was dictated by the protocol was performed, the ongoing pregnancy rate was 20% (25/128) with a multiple pregnancy rate of 0% (0/128) ($p=0.004$), and in those where TET was performed, the corresponding rates were 19% (5/26) and 40% (2/5). In the cycles of women with a poor prognosis the ongoing pregnancy rate was 0% (0/6) after SET, and 20% (25/126) after DET with a multiple pregnancy rate of 8% (2/25).

DISCUSSION

One of the major unsolved issues in modern IVF is optimizing pregnancy rates while limiting multiple pregnancy rates in women with an intermediate or poor prognosis, such as women over 35 years of age with one or more failed IVF cycles. This prospective evaluation of a tailored embryo transfer policy based on the prognostic profile of the individual patient, shows that in women with an intermediate prognosis double embryo transfer results in good ongoing pregnancy rates, but at the expense of high multiple pregnancy rates. Women with a poor prognosis undergoing triple embryo transfer have reasonable ongoing pregnancy rates with a limited multiple pregnancy rate of 13%.

The use of SET in women with a good prognosis resulted in high ongoing pregnancy rates with very low multiple pregnancy rates. In 87 cycles with protocol violation in which two top-quality embryos were transferred instead of one, similar pregnancy rates but significantly higher multiple pregnancy rates were obtained compared to SET. This again confirms that SET should be the preferred strategy in women with a good prognosis. Women with a good prognosis have a very good chance of pregnancy, regardless of the number of embryos transferred; transferring multiple embryos would only result in more multiple pregnancies, not in higher pregnancy rates.

In women with an intermediate prognosis ongoing pregnancy rates were good (cumulative 27%), but multiple pregnancy rates were too high (cumulative 23%). In these women better selection is obviously necessary. A uniform change of transfer policy for all women with an intermediate prognosis from DET to eSET would obviously reduce multiple pregnancy rates, but extrapolating the data from two meta-analyses from synthesized aggregated data and individual patient data on eSET versus DET in women with a good prognosis, it may well be that pregnancy rates would also be reduced (8;9). Expanding the use of an eSET policy to older women but who react well to the IVF procedure in terms of
good quality embryos, might be the way to go to reach good pregnancy rates with lower multiple pregnancy rates. In a previous cohort study eSET was offered to a selected group of women with a less favourable prognosis based upon age (women 36-39 years), but who responded well to ovarian stimulation, had at least one good quality embryo for transfer and at least one additional embryo that could be frozen (13). The pregnancy rates after DET were similar to those in our study and eSET resulted in similar pregnancy rates but lower multiple pregnancy rates than DET.

In women with a poor prognosis triple embryo transfer resulted in reasonable ongoing pregnancy rates and a multiple pregnancy rate of 13%. One recent large cohort study also evaluated SET, DET and TET in women with a poor prognosis, i.e. women over 39 (24). TET in women with a poor prognosis resulted in similar pregnancy rates compared to DET but also in higher multiple pregnancy rates (24). This is in contrast to the results from a large Society of Assisted Reproductive Technology database study evaluating the optimal number of cleavage stage embryos to transfer in women with a poor prognosis, i.e. women of 38 years or older (25). They found that pregnancy rates increased up to the transfer of three embryos in 38 year old women and even up to four embryos in 39 year old women (25). A possible explanation for these discrepancies between these studies is selection bias. Since there was no strict embryo transfer policy in the first study, it is plausible that women with a poorer prognosis received three embryos instead of two embryos. In our study we used a strict transfer policy, therefore minimizing selection bias. Our findings are therefore probably more robust, and show that transferring three embryos still has a place in this accurately defined subgroup of women.

An important question remains: what are acceptable multiple pregnancy rates in women with a poor prognosis? Most couples desire more than one child. When determining the risk/benefit of the number of embryos to be transferred, the number of desired children should also be taken into consideration (26). The time required for multiple frozen cycles in a SET strategy or to initiate a new cycle after a failed IVF/ICSI cycle may result in additional loss of fertility potential. In young women with a good prognosis who have the highest chance of a twin pregnancy, this delay may not be clinically relevant. Yet, for women of older reproductive age this is a relevant concern. For these women a twin pregnancy may be the only chance to conceive and deliver a second child and many studies have shown that a significant proportion of women have a clear preference for a twin pregnancy, regardless of any intervention trying to influence this preference (27-32). So, when deciding how many embryos to transfer, the focus should not be solely on health care costs associated with twin pregnancies, but should equally take into account, after proper counselling on short and long term adverse perinatal outcomes of twin pregnancy, the preference of the couples undergoing treatment.
Chapter 6

Pregnancy rates after frozen-thawed embryos are increasing; in the near future these might almost reach the pregnancy rates achieved after a fresh embryo transfer. If so, the reduction in pregnancy rates from fresh SET could be remedied by subsequent SET in a frozen-thawed cycle. Awaiting randomised embryo transfer trials that evaluate all frozen-thawed embryo transfers from the same cycle, a more individualized embryo transfer strategy - one that takes the whole prognostic profile of women into account - could be a way to optimize pregnancy rates at a low risk of unwanted multiple pregnancies. Several studies attempted to refine transfer policy by developing such prediction models (33-35). Unfortunately, these models were developed for the first IVF cycle only, focused mainly on single- or double embryo transfer and did not evaluate embryo quality extensively, i.e. they did not include important embryo predictors such as early cleavage or number of blastomeres as a separate predictor. Refining and implementing such models could be the way to achieve optimal pregnancy rates at a low risk of unwanted multiple pregnancies. In the meantime, the data in this study can be used to guide current practice, i.e. performing SET in women with a good prognosis and TET in women with a poor prognosis. The embryo transfer strategy in women with an intermediate prognosis requires further improvement possibly by refining the transfer strategy according to the ovarian response after ovarian stimulation.
Pregnancy and twinning rates using a tailored embryo transfer policy.

REFERENCES


Pregnancy and twinning rates using a tailored embryo transfer policy.


Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential using morphological scoring

L.L. van Loendersloot
M. van Wely
S. Repping
P.M.M. Bossuyt
F. van der Veen

Reproductive BioMedicine Online conditionally accepted September 2013
Selection of embryos based on morphology is still the core of daily laboratory practice in IVF/ICSI. Yet, the selection of embryos is primarily based on experience and local protocols. Since an evidence-based ranking strategy for embryos on day 3 is currently lacking, we constructed a multivariable prediction model, to rank embryos according to their implantation potential. We studied 6,021 fresh embryo transfers between January 2004 and July 2009. We evaluated pronuclear score, early cleavage, number of blastomeres on day 2/3, morphological score on day 2/3 and cleavage rate as potential predictors for ongoing implantation. The outcome measure was ongoing implantation. A model was developed using multivariable logistic regression. This prediction model was externally validated with data from couples treated between August 2009 and September 2011 in the same clinic. Five factors were included in the final prediction model. In the external validation the model showed moderate discriminative capacity (c-statistic 0.70) and calibrated well. The model was able to distinguish embryos with high ongoing implantation potential from embryos with moderate or low ongoing implantation potential. The model can be used by embryologists as an objective tool to rank embryos according to their implantation potential thereby aiding the selection of embryos for transfer.
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

INTRODUCTION

In the early days of in vitro fertilization (IVF), pregnancy rates per embryo transfer were very low (1). The only way to increase pregnancy rates at that time was the transfer of large numbers of embryos (2). Over the years implantation rates per embryo improved and the transfer of these large numbers of embryos led to high multiple pregnancy rates (3-5). To reduce these high multiple pregnancy rates, embryo transfer policies restricting the number of embryos to be transferred were introduced. This meant that selection of the embryo(s) with the highest chance of implantation became of paramount importance in IVF, especially since the success rates of cryopreservation of supernumerary embryos were still low at the start of cryopreservation programs (3).

In the last decade, cryopreservation of embryos has become increasingly successful and this has placed embryo selection in a new context (3;6). Optimal selection of embryos can help to minimize the time to pregnancy by transferring the embryo with the highest implantation potential as early as possible. Transferring only one embryo if there is a high chance of implantation could also help to achieve acceptable pregnancy rates while minimizing the chances of multiple pregnancy.

Ever since the start of IVF the selection of embryos has been largely based on morphological characteristics of the embryo. Additional methods for embryo selection, such as selection based on chromosomal status (PGS) or selection based on metabolomic profiles of culture media, have been introduced but upon proper evaluation these methods have been shown to be unable to increase pregnancy rates (7-9). Morphological selection of embryos thus remains the core of daily laboratory practice in IVF/ICSI. Yet, morphological selection of embryos is largely based on clinical experience and local protocols (10;11).

Several authors proposed prediction models to rank embryos according to their implantation potential. Unfortunately, most of these models were developed on small datasets and were not externally validated (11-18). Based on data from a large cohort of consecutively treated IVF/ICSI patients, we constructed a new multivariable prediction model, to rank embryos according to their implantation potential.

MATERIALS AND METHODS

We collected data consecutive IVF/ICSI embryo transfers on day 3 after oocyte retrieval performed between January 2004 and July 2009 in the Center for Reproductive Medicine of the Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands, for the development of the model.
For validation of the model we prospectively collected data from embryo transfers performed between August 2009 and September 2011 at the same center. Under the legal requirements for clinical research in the Netherlands, this study was exempt from institutional review board (IRB) approval.

Patients
All couples in our study had been trying to conceive for at least 12 months and underwent a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynecology (19). The indication to start IVF or ICSI treatment was determined according to the Dutch IVF guideline (20). If subfertility was caused by tubal pathology, severe endometriosis, or severe oligozoospermia (post-wash total motile sperm count < 3 million) IVF/ICSI was offered directly (21). In case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia, or unexplained subfertility, at least six intrauterine inseminations (IUI) were applied before IVF/ICSI was offered. In case of ovulation disorders, mainly caused by polycystic ovary syndrome (PCOS), 12 cycles of ovulation induction were applied before IVF/ICSI was offered.

IVF/ICSI procedures
Women underwent controlled ovarian hyperstimulation after down-regulation with the GnRH agonist triptorelin (Decapeptyl®) in a long protocol with a midluteal start. Controlled ovarian hyperstimulation was started on cycle day 5 with recombinant FSH or HMG in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Follicular maturation was induced by 10,000 IU human chorionic gonadotropin hormone (hCG) (Pregnyl, Organon). Cumulus-oocyte complexes were recovered by transvaginal ultrasound guided follicle aspiration 36 hours thereafter. Oocytes were inseminated with 10,000 or 15,000 progressively motile spermatozoa (in vitro fertilization) or injected with a single spermatozoon (intracytoplasmic sperm injection) 2-4 hours after follicle aspiration. Embryos were cultured in Human Tubal Fluid (HTF, Cambrex) supplemented with 15% pasteurized plasma protein solution (GPO, Sanquin) or G5-PLUS medium (Vitrolife) containing Human Serum Albumin at 37°C and 5% (HTF) or 6% (G5) CO₂ in air. Embryo transfer was performed on day 3 after oocyte retrieval with a Wallace catheter (Smiths Medical). Luteal phase was supported by progesterone intravaginally two times 200 mg (Utrogestan) per day. An hCG blood test was performed 18 days after oocyte retrieval.

Morphological scoring
Each embryo was cultured individually. Pronuclear scoring was performed 17 to 22 hrs. after insemination/injection and early cleavage was scored 23 to 28 hours after insemination/injection. On day 2 (41 to 46 hrs. after insemination/injection) and day 3 (65 to 70 hrs. after insemination/injection) the number of blastomeres was assessed and each embryo was given a morphological score. For the morphological score the degree of fragmentation of the embryo and the uniformity of the blastomeres were assessed (22). The embryo was given a score of 1 (no fragmentation), 2 (<20% fragmentation), 3 (20-50%
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

Fragmentation or 4 (>50% fragmentation). If the blastomeres of the embryo were non-uniform in size for their developmental stage (i.e., the 2, 4, or 8 cell stage), the morphological score was augmented with one point with 4 remaining the lowest possible score. If on day 3 the embryo showed signs of compaction, the embryo was scored as a morula and given a grade based on the degree of compaction (score 1: full compaction, score 2: 50 - <100% compaction and score 3: less than 50% compaction).

Number of embryos transferred
Before July 2006 in all women double embryo transfer was performed unless there was a medical indication to limit the number of transferred embryos to one. After July 2006, an individualized transfer policy was adopted. Single embryo transfer (SET) was performed in women under 35 years undergoing their first cycle of IVF/ICSI with at least one top-quality embryo. Double embryo transfer (DET) was performed in women under the age of 35 who did not have a top quality embryo in the first cycle, in women under 35 who failed to get pregnant in their first cycle of IVF/ICSI, and in women between 35 and 38. In women 39 or older three embryos were transferred (TET). These strategies were based on a combination of the Practice Committee of the American Society for Reproductive Medicine guidelines and the Belgian embryo transfer legislation (23;24).

Outcome
The primary outcome was ongoing implantation, defined as the presence of one fetus with cardiac activity at transvaginal ultrasound at a gestational age of at least 11 weeks per embryo transferred.

Predictors
We evaluated pronuclear score, early cleavage, number of blastomeres on day 2 and day 3, morphological score on day 2 and day 3 and the progression of the number of blastomeres from day 2 to day 3 as potential predictors for ongoing implantation. Since we wanted to develop a model to rank embryos, not to calculate the chances of success, we only evaluated embryo parameters, leaving out female and male characteristics. Since within each individual cycle couple and treatment characteristics will be identical for all embryos, they are of no help in ranking embryos according to their implantation potential.

Data analysis
For the development of the model, only cycles with embryos with individual traceability were used. These were cycles with single-, double- or triple embryo transfer resulting in either no implantation or transplantation of all transferred embryos. Monozygotic twins were excluded from the analysis. Embryos on which PGS was performed were also excluded. The embryo was the unit of analysis in model development. This dataset is the development dataset with traceable embryos.
Chapter 7

Some of the candidate predictors had missing values. Simple exclusion of couples with missing values on one or more variables commonly causes biased results and decreases statistical efficiency (25). Therefore, missing values in the data were completed by multiple imputation using SPSS (version 18.0). This method uses all available data to impute the missing values based on the correlation between each variable with missing values and all other variables. If a single candidate predictors had more than 25% missing values, they were excluded from the analyses.

We checked the linearity of the associations between the continuous variables number of blastomeres and the categorical variable morphological score and the probability of an ongoing implantation using restricted cubic spline functions in logistic regression and visual inspection. Based on these spline functions, variables were transformed to better approach linearity.

For each candidate predictor, we performed a univariable logistic regression analysis and estimated the corresponding unconditional odds ratio, calculating 95% confidence intervals (CI) and the p-value. All predictors that were significantly associated with ongoing implantation (p<0.3) were entered in a multivariable logistic regression analysis. In deciding between competing expressions of related parameters, we used Akaike’s Information Criterion (AIC) in variable selection (26). The model with the best AIC was selected as the final model. Additionally, we evaluated all potential predictive factors for interactions using an interaction term.

To prevent overfitting and a too optimistic impression of model performance, a linear shrinkage factor was estimated based on model fit and the number of parameters (26). Coefficients in the model were then corrected by this shrinkage factor.

**Performance**

Performance of the final model was first evaluated by assessing the ability of the model to distinguish between embryos that achieved an ongoing implantation and those that did not. To evaluate discrimination of the model, the area under the receiver operating characteristic curve (AUC), also known as the c-statistic, was calculated. The c-statistic expresses the probability that in any pair of embryos, in which one implanted and the other did not, the embryo that implanted actually had a higher score.

To extrapolate the implantation rates from the dataset with individual embryo traceability to the total dataset a correction factor was used. The correction factor was calculated as the ratio of the overall implantation rate – the number of implantations relative to the number of transferred embryos in the total dataset – versus the implantation rate in the individual traceability dataset; the number of implantations relative to the number of transferred embryos in the individual traceability dataset.
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

We calculated for each traceable embryo in development dataset and in the validation dataset the implantation probability. Ideally, these probabilities should show a wide range, making it easier to rank embryos based on their implantation potential.

To evaluate agreement between calculated probabilities of an ongoing implantation and observed proportions of ongoing implantation, we calculated the Hosmer-Lemeshow goodness-of-fit test statistic.

In addition we compared the average calculated probabilities of ongoing implantation in disjoint subgroups defined by quintiles with the observed ongoing implantation rate in the corresponding groups. The predicted proportion and the observed proportion of ongoing implantations (for traceable embryos) were compared by plotting the observed ongoing implantation rate versus the average probability in each of the groups, as calculated from the model.

To evaluate any miscalibration, we also fitted a calibration model using logistic regression, with the linear combination of variables in the prediction model as the only variable (26;27).

**External validation**

We performed an external, temporal validation (26). We evaluated model performance in more recent embryo transfers on day 3 after oocyte retrieval, performed between August 2009 and September 2011 in the Centre for Reproductive Medicine of the Academic Medical Centre, the Netherlands. We performed the validation on the validation dataset with traceable embryos. We also performed a validation on the complete development set and validation set (on a transfer level) containing all transferred embryos. The probability of success in transfer was calculated based on the calculated probabilities of the embryos transferred. The predicted proportion and the observed proportion of success (for all transfers) were compared by plotting the observed ongoing implantation rate versus the average probability in each of the groups, as calculated from the model.

**Updating the model**

After the external validation, we updated our model based on all available data through re-calibration (28;29). We fitted the linear combination of the variables in our model as the only variable in a logistic regression model, using all traceable embryos in the development set and the validation set. Based on the estimated slope and intercept of that model we adjusted the intercept and coefficients of our prediction model, to create a final, updated model.
Chapter 7

RESULTS

Between January 2004 and July 2009, 3,143 embryo transfers had been performed, transferring a total of 6,021 embryos (average 1.9 embryos per transfer). Of these, 848 implanted: a total ongoing implantation rate of 14.1%. The transfers led to a viable intra-uterine pregnancy of at least 11 weeks in 713 cases (23%); in 247 transfers all embryos transferred implanted, in 466 transfers fewer embryos implanted than transferred, and in 2,430 embryo transfers no embryos implanted (Suppl Fig 1). A total of 5,028 embryos had exact traceability and were used further for model development.

Between August 2009 and September 2011, 1,666 additional embryo transfers were performed, transferring a total of 3,060 embryos (average 1.8 embryos per transfer). The ongoing pregnancy rate in this validation set was 21% (351/1,666). In 152 transfers all embryos transferred implanted, in 199 transfers less embryos implanted than transferred, and in 1315 embryo transfers no embryos implanted (Suppl Fig 1).

The baseline characteristics of all embryo transfers and the datasets with traceable embryos are summarized in Suppl Table I, both for the development set and for the validation set.

Analysis with spline functions demonstrated a nonlinear association between the number of blastomeres on day 2 and day 3 after oocyte retrieval and ongoing implantation. We transformed both variables to better fit the data. The number of blastomeres on day 2 was recoded as the absolute value of the deviation from 4; an embryo with six blastomeres was recoded to a score two (6 minus 4) and an embryo with three blastomeres was recoded to a score of 1 (4 minus 3). Similarly, the number of blastomeres on day 3 was recoded as the absolute value of the deviation from 8. All embryo morphology scores could adequately be described by linear functions (Suppl Fig 2).

In univariable analysis, early cleavage, number of blastomeres on day 2 and 3, morphological score on day 2 and day 3, and progression from 4 blastomeres on day 2 to 8 blastomeres on day 3 were found to be significantly associated with ongoing implantation. The pronuclear score (p=0.26) and the presence of morulae on day 3 (p=0.2) were moderately associated with ongoing implantation (Suppl Table 2).

In multivariable analysis five factors were found to be significantly associated with ongoing implantation: early cleavage, number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morulae on day 3. These factors were included in the final multivariable logistic regression model (Table I). None of the evaluation interactions between these terms was statistically significant; no interaction terms were included in the final model.

We compared the goodness-of-fit of the final model to that of two other models: a model with only day 3 blastomeres and a model with day 3 blastomeres and the morphological
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential score on day 3. Our final model fitted the data significantly better than the other two models (likelihood ratio test; \( p < 0.001 \))

Figure 1 depicts the spread of predicted probabilities in the development set and validation set with traceable embryos. The calculated probabilities for both datasets ranged from 0.00 to 0.39, with a mean of 0.14, which corresponds to the overall implantation rate in this selected set of embryos.

Figure 1 | Distribution of the calculated probabilities in the development set and validation set with traceable embryos

The model had a moderate discriminative capacity in the development set with traceable embryos. The c-statistic was 0.73 (95% CI: 0.70 to 0.75). There was good calibration; the goodness-of-fit test (Hosmer-Lemeshow) showed no significant miscalibration (\( p=0.27 \)), the calibration slope was 1.02 (95% CI: 0.87 to 1.18) and the calibration intercept was 0.05 (95% CI: -0.31 to 0.42). The calibration plot showed that the model calibrated well (Figure 2a).

Discriminative capacity in the validation set with traceable embryos was similar to that in the development set, with a c-statistic of 0.70 (95% CI: 0.67 to 0.74). In the validation set the model also calibrated well (Figure 2b). The slope of the linear predictor (calibration slope) was 0.89 (95% CI: 0.69 to 1.09); the calibration intercept was -0.26 (95% CI: -0.74 to 0.24).

We also evaluated performance of the model for all embryo transfers. The model calibrated well both in the complete development set and validation set (Figure 2c and d).
Chapter 7

Figure 2 | Calibration plots, showing the association between the calculated and observed embryo ongoing implantation rates

a. Development set with traceable embryos

b. Validation set with traceable embryos

c. Validation complete development set

d. Validation complete validation set

The updated final model and a simplified embryo score are presented in Table I. The total score can be calculated with the following formula:

Total score = 103 + (2 * early cleavage (yes=1, no=0)) + (-3 * number of blastomeres on day 2 deviating from 4) + (-3 * number of blastomeres on day 3 deviating from 8 (morula=0)) + (-5 * morphological score on day 3 (morula=0)) + (-11 * morula (yes=1, no=0)).
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

Table I | Multivariable analysis

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Updated model</th>
<th>Embryo Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.0579</td>
<td>103</td>
</tr>
<tr>
<td>Early cleavage</td>
<td>0.2492</td>
<td>2</td>
</tr>
<tr>
<td>Number of blastomeres on day 2 calculated as: value of the deviation from 4</td>
<td>-0.3324</td>
<td>-3</td>
</tr>
<tr>
<td>Number of blastomeres on day 3 calculated as: value of the deviation from 8</td>
<td>-0.3128</td>
<td>-3</td>
</tr>
<tr>
<td>Morphological score day 3</td>
<td>-0.5305</td>
<td>-5</td>
</tr>
<tr>
<td>Morula on day 3</td>
<td>-1.1940</td>
<td>-11</td>
</tr>
</tbody>
</table>

1 Number of blastomeres = absolute value (number of blastomeres - 4)  
2 Number of blastomeres = absolute value (number of blastomeres - 8), morula = 0  
3 Morula = 0  
4 Presence of morula=1, no morula = 0.

The higher the total score, the higher the ongoing implantation potential of the embryo.

Table II depicts a hypothetical case of a couple that has 10 embryos after an IVF/ICSI cycle. Their embryos are ranked according to their implantation potential based on our model. For example, an embryo with early cleavage, 4-blastomeres on day 2, 8-blastomeres on day 3 with a morphological score of 1 on day 3 would have a total score of 100 (103 + (2*1) + (-3*0) + (-3*0) + (-5*1) + (-11*0)) (Table II, embryo 10). The embryo with the highest total score has the highest chance of implantation compared to the other 9 embryos.

Table II | Hypothetical example of the ranking of embryos three days after oocyte retrieval

<table>
<thead>
<tr>
<th></th>
<th>Early cleavage</th>
<th>Number of blastomeres day 2†</th>
<th>Number of blastomeres day 3‡</th>
<th>Morphological score day 3</th>
<th>Morula on day 3</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo 1</td>
<td>no</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>no</td>
<td>67</td>
</tr>
<tr>
<td>Embryo 2</td>
<td>no</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>no</td>
<td>76</td>
</tr>
<tr>
<td>Embryo 3</td>
<td>no</td>
<td>4</td>
<td>12</td>
<td>2</td>
<td>no</td>
<td>81</td>
</tr>
<tr>
<td>Embryo 4</td>
<td>no</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>no</td>
<td>87</td>
</tr>
<tr>
<td>Embryo 5</td>
<td>no</td>
<td>3</td>
<td>NA</td>
<td>NA</td>
<td>yes</td>
<td>89</td>
</tr>
<tr>
<td>Embryo 6</td>
<td>yes</td>
<td>4</td>
<td>9</td>
<td>2</td>
<td>no</td>
<td>92</td>
</tr>
<tr>
<td>Embryo 7</td>
<td>no</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>no</td>
<td>93</td>
</tr>
<tr>
<td>Embryo 8</td>
<td>yes</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>yes</td>
<td>94</td>
</tr>
<tr>
<td>Embryo 9</td>
<td>no</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>no</td>
<td>98</td>
</tr>
<tr>
<td>Embryo 10</td>
<td>yes</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>no</td>
<td>100</td>
</tr>
</tbody>
</table>

† The number of blastomeres are transformed to: abs (number of blastomeres - 4)  
‡ The number of blastomeres are transformed to: abs (number of blastomeres - 8), morula= 0  
NA: Not applicable
Chapter 7

DISCUSSION

In this study we developed a prediction model to rank embryos within a single IVF/ICSI cycle of a couple according to their ongoing implantation potential. The model had moderate discriminative capacity and calibrated well, both in the development and in a separate validation set, with data that had not been used for the development of the model.

One of the strengths of our study is that we evaluated seven embryo predictors in consecutive transfers, using only embryos with exact traceability. We developed our model in a large data set (>6,000 embryos) and we validated our model thoroughly using more recent data, collected at the same clinic after the development of the model.

The moderate discriminative capacity implies that the model is not able to distinguish perfectly between embryos with small differences in ongoing implantation potential. Yet perfect prediction is not the goal of this embryo selection model: the primary goal is not to predict with absolute certainty whether an embryo will implant, but to rank embryos based on their ongoing implantation potential within a single treatment cycle of a couple. Although the ideal outcome of the model would be the number of live births, our study used ongoing implantation rate as the outcome of interest. We have not yet evaluated whether implementation of this model ultimately improves ongoing implantation rates and time to pregnancy, a topic for future research. Since < 2% of all ongoing pregnancies result in late miscarriage or still birth, we do not expect that our model will fundamentally change when using the number of live births as outcome measure (30).

When scoring the embryos, there was a range in timing of scoring of maximally 5 hours. This range could potentially lead to different classification of embryos. Yet despite the range in timing of scoring the embryos, the model had an acceptable discriminative capacity and calibrated perfectly even after external validation. We used data of a single center only, so the generalizability of the model to other clinics has to be evaluated more extensively in future studies (geographical validation). As the aim of the model was to rank the embryos acquired after an IVF/ICSI cycle of a couple and not to calculate the exact implantation rate of an individual embryo, higher or lower implantation rates should not influence the performance of the model. Over the years the data in this study were collected, there were two significant changes in our center: the embryo transfer policy shifted more towards single embryo transfer (as seen by the lower number of embryos transferred in the validation set), and a switch in culture media was made (HTF to Vitrolife). Despite these changes our model still had near-perfect calibration and acceptable discriminative capacity both after internal and external validation, indicating that these changes did not affect the performance of the model.
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

As indicated in the introduction, several other embryo implantation models have been developed in the past. Previous studies used much smaller datasets and not all used data of embryos with exact traceability of the individual embryos (11-17). Several studies did not validate their model (11-14;17). As prediction models may not perform as well in a new dataset as in the development set, external validation of models is essential to support general applicability of the model (26). It also enables further fine tuning of the model by updating the weight of each variable.

An additional problem with some other embryo implantation models is that they included patient characteristics into the model (13;18). As patient characteristics for each of these embryos are identical - all embryos are from the same couple - it is misleading to include these characteristics. They will seem to improve model fit, without actually contributing to the ranking potential. Some of these patient characteristics, such as age, are much stronger predictors than embryo parameters, so including these in a model would result in an overestimation of the discriminative capacity of the model in distinguishing between embryos of the same women, in which female age is identical.

The association between the five identified embryo factors and embryo implantation is biologically plausible: embryos that demonstrate early cleavage are known to be more likely to implant because they are likely to cleave more evenly, which is strongly correlated with a lower incidence of mitotic chromosomal errors, and therefore a higher chance of implantation (31). Our analyses showed that faster and slower cleaving embryos have a lower chance of implantation. The biological explanation could be that embryos that cleave directly into more than two cells and embryos that cleave too slowly or too fast also have significantly more chromosomal abnormalities (32;33). In addition, embryos that cleave faster have recently been shown to have greater perturbations in genomic imprinting and metabolic marker expression (34). The degree of fragmentation of an embryo is strongly correlated with chromosomal mosaicism and embryos that display fragmentation are less likely to implant (35).

Early cleavage (day 1) and the number of blastomeres on day 2 are important predictors in our model, implying that embryo selection should not be solely based on embryo parameters assessed on day 3. Culturing embryos individually and scoring them on each day therefore allows for better embryo selection. Newly developed real-time embryo monitoring systems enable the continuous monitoring of embryos and could assist in accurate determination of the timing of all cleavages (36). In the absence of sufficiently large randomized clinical studies it remains to be elucidated whether embryo selection using dynamic parameters improves clinical outcome, or whether it has additional predictive capacity for implantation. Before such randomized controlled studies are performed morphological selection based on daily evaluation of the embryos seems to remain at the core of current laboratory practice in IVF/ICSI.
Combining a multivariable model that takes both the prognostic profile of the patient and the ranking of the embryos into account could be an important step towards a ‘patient tailored’ embryo transfer strategy. Such a combined model would enable calculation of ongoing implantation chances and multiple ongoing implantation chances, This is especially relevant in the situation where a decision has to be made to transfer one or two embryos. Currently, embryo quality is mostly dichotomized into top quality and non-top quality embryos. Our study shows that embryos can be ranked more precisely, based on their ongoing implantation potential, and that dichotomizing embryo quality is most likely an oversimplification of reality.

In the meantime, the model presented here can be used by embryologists as an objective tool to rank embryos by their ongoing implantation potential, and to select the embryo(s) with the highest ongoing implantation potential for transfer. Furthermore, it can help to create a more uniform embryo selection strategy for all laboratories transferring embryos on day 3 after oocyte retrieval.
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

REFERENCES


Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential


Supplemental figure 1 | Flowchart on embryo transfers

1a. Development dataset

1b. Validation dataset
Supplemental figure 2 | Spline functions that visualizes the association of the prediction of ongoing implantation and embryo variables

- A. Number of blastomeres on day 2
- B. Morphological score on day 2
- C. Number of blastomeres on day 3
- D. Morphological score on day 3
## Supplemental table I | Baseline characteristics development and validation set

<table>
<thead>
<tr>
<th></th>
<th>Development set</th>
<th>Development set traceable embryos</th>
<th>Validation set</th>
<th>Validation set traceable embryos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Number of patients</td>
<td>1790</td>
<td>1569</td>
<td>1114</td>
<td>%</td>
</tr>
<tr>
<td>Number of IVF cycles</td>
<td>3143</td>
<td>2677</td>
<td>1666</td>
<td>%</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>35.5 (4.5)</td>
<td>35.6 (4.5)</td>
<td>35.9 (4.5)</td>
<td>35.8 (4.5)</td>
</tr>
<tr>
<td>Indication for IVF/ICSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal pathology</td>
<td>616</td>
<td>20%</td>
<td>538</td>
<td>20%</td>
</tr>
<tr>
<td>Unexplained subfertility</td>
<td>875</td>
<td>28%</td>
<td>739</td>
<td>28%</td>
</tr>
<tr>
<td>Male subfertility</td>
<td>1395</td>
<td>44%</td>
<td>1168</td>
<td>44%</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>99</td>
<td>3%</td>
<td>86</td>
<td>3%</td>
</tr>
<tr>
<td>Oocyte donation</td>
<td>24</td>
<td>1%</td>
<td>24</td>
<td>1%</td>
</tr>
<tr>
<td>Others</td>
<td>380</td>
<td>12%</td>
<td>336</td>
<td>13%</td>
</tr>
<tr>
<td>Number of embryo transfers</td>
<td>3143</td>
<td>2677</td>
<td>1666</td>
<td>1467</td>
</tr>
<tr>
<td>Number of embryos transferred</td>
<td>6021</td>
<td>5028</td>
<td>3061</td>
<td>2617</td>
</tr>
<tr>
<td>Number of single embryo transfers</td>
<td>605</td>
<td>605</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Number of double embryo transfers</td>
<td>2198</td>
<td>1793</td>
<td>837</td>
<td>683</td>
</tr>
<tr>
<td>Number of triple embryo transfers</td>
<td>340</td>
<td>279</td>
<td>279</td>
<td>234</td>
</tr>
<tr>
<td>Number of ongoing pregnancies</td>
<td>713</td>
<td>247</td>
<td>351</td>
<td>152</td>
</tr>
<tr>
<td>Singletons</td>
<td>578</td>
<td>120</td>
<td>297</td>
<td>105</td>
</tr>
<tr>
<td>Twins</td>
<td>135</td>
<td>127</td>
<td>54</td>
<td>47</td>
</tr>
<tr>
<td>Triplets</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of ongoing implantations</td>
<td>848</td>
<td>374</td>
<td>405</td>
<td>199</td>
</tr>
</tbody>
</table>
Selection of embryos for transfer in IVF: ranking embryos based on their implantation potential

**Supplemental table II | Univariable analysis**

<table>
<thead>
<tr>
<th>Univariable analysis</th>
<th>( \beta )</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronuclear score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 PN (reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 PN</td>
<td>-1.25</td>
<td>0.28</td>
<td>(0.07 - 1.17)</td>
<td>0.08</td>
</tr>
<tr>
<td>0 PN</td>
<td>-0.18</td>
<td>0.84</td>
<td>(0.55 - 1.28)</td>
<td>0.41</td>
</tr>
<tr>
<td>Unknown</td>
<td>-1.15</td>
<td>0.32</td>
<td>(0.04 - 2.32)</td>
<td>0.26</td>
</tr>
<tr>
<td>2-blastomeres</td>
<td>0.48</td>
<td>1.61</td>
<td>(0.37 - 7.07)</td>
<td>0.53</td>
</tr>
<tr>
<td>Early cleavage</td>
<td>0.65</td>
<td>1.92</td>
<td>(1.42 - 2.59)</td>
<td>0.00</td>
</tr>
<tr>
<td>Number of blastomeres day 2</td>
<td>0.24</td>
<td>1.28</td>
<td>(1.15 - 1.41)</td>
<td>0.00</td>
</tr>
<tr>
<td>Number of blastomeres day 3</td>
<td>0.32</td>
<td>1.37</td>
<td>(1.28 - 1.47)</td>
<td>0.00</td>
</tr>
<tr>
<td>Morulae</td>
<td>0.32</td>
<td>1.37</td>
<td>(0.84 - 2.23)</td>
<td>0.20</td>
</tr>
<tr>
<td>Morphological score day 2</td>
<td>-0.83</td>
<td>0.04</td>
<td>(0.35 - 0.54)</td>
<td>0.00</td>
</tr>
<tr>
<td>Morphological score day 2</td>
<td>1</td>
<td>Reference</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.592</td>
<td>0.55</td>
<td>(0.40 - 0.4)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.724</td>
<td>0.18</td>
<td>(0.18 - 0.11)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphological score day 3*</td>
<td>-1.05</td>
<td>0.35</td>
<td>(0.28 - 0.43)</td>
<td>0.00</td>
</tr>
<tr>
<td>Morphological score day 3*</td>
<td>1</td>
<td>Reference</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.85</td>
<td>0.43</td>
<td>(0.32 - 0.57)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.23</td>
<td>0.10</td>
<td>(0.06 - 0.16)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progression from 4 blastomeres on day 2</td>
<td>1.28</td>
<td>3.59</td>
<td>(2.89 - 4.45)</td>
<td>0.00</td>
</tr>
<tr>
<td>to 8 blastomeres on day 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* for analysis morula embryos were excluded

OR= odds ratio
CI= confidence interval
\( \beta \)= beta coefficient
Balancing pregnancy and multiple pregnancy chances: an embryo transfer model

L. L. van Loendersloot
M. van Wely
F. van der Veen
S. Repping
P.M.M. Bossuyt

Manuscript in preparation
Chapter 8

ABSTRACT

Background
In women with a good prognosis eSET provides a well-balanced tradeoff between pregnancy and multiple pregnancy. Many women currently undergoing IVF have an intermediate or poor prognosis. In these women the critical question is how many embryos to transfer to obtain optimal pregnancy rates while simultaneously limiting the risk of multiple pregnancy. To aid clinical decision making in these women, we developed a multivariable embryo transfer model that calculates the probability of a pregnancy as well as the chances of multiple pregnancy after single, double embryo or triple embryo transfer.

Methods
The embryo transfer model consists of two components: variables specific to the transferred embryo(s), such as the number of blastomeres and variables specific to the couple, including maternal and treatment factors that affect all embryos equally. We combined two previously described models, one for ranking embryos based on implantation potential and one for estimating pregnancy chances. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic curve (AUC) and calibration.

Results
In total, 2,201 IVF/ICSI cycles in 1297 couples could be included in our analyses. The embryo transfer model includes the following variables specific for the couple: female age, duration of subfertility, previous ongoing pregnancy, male subfertility diminished ovarian reserve, endometriosis, basal FSH and number of failed IVF cycles and in case of a failed IVF cycle: presence of absence fertilization, number of embryos, mean morphological score, presence of 8-cell embryos and presence of morulae on Day 3 of the previous cycle, and the following variables specific for the actual IVF cycle: early cleavage, number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morulae on day 3. The model calibrated well. The AUC for predicting pregnancy was 0.692 and multiple pregnancy was relatively high, at 0.778

Conclusion
We have developed an embryo transfer model to calculate the probability of pregnancy and multiple pregnancy following single, double or triple embryo transfer. Our model may aid in the decision how many embryos to transfer in any specific woman with any specific embryos to prevent a multiple pregnancy without reducing the chances of a pregnancy. Wider applicability of this model requires external validation in databases from other centers.
INTRODUCTION

In the early days of in vitro fertilization (IVF), pregnancy rates per embryo transfer were very low (1). The only way to increase pregnancy rates at that time was the transfer of large numbers of embryos (2). Over the years implantation rates per embryo improved and the transfer of these large numbers of embryos led to high multiple pregnancy rates (3-5). Initially, a multiple pregnancy was felt to be justified in what was largely an experimental treatment, with poor pregnancy and live birth rates, but in the last two decades pregnancy rates have increased substantially but multiple pregnancy rates have remained high (3;6;7). In 2009, the pregnancy rate in Europe was 19.7% and 20.2% of these pregnancies were multiple pregnancies (8). These high multiple pregnancy rates caused concern, since multiple pregnancy is associated with increased maternal and perinatal morbidity and mortality, as well as with increased health care costs (9-12).

To curtail the multiple birth risk, treatment strategies reducing the number of embryos transferred were introduced. The first strategy was to reduce triple embryo transfers and to increase double embryo transfers. This indeed prevented most triplets, but did not diminish the rate of twin pregnancies (7). Later, elective single embryo transfer (eSET) was introduced. In women with a good prognosis, eSET largely prevents multiple pregnancies, compared to double embryo transfer (DET), but it also halves the odds of a live birth per fresh cycle. Cryopreservation of supernumerary embryos from the same eSET cycle and subsequent transfer(s) of a single frozen thawed embryo results in cumulative live birth rates comparable to those after double embryo transfer (13;14).

Although eSET is now an accepted policy for women with a good prognosis – those younger than 36 years undergoing a first or second cycle with at least two good quality embryos – many women currently undergoing IVF have an intermediate or poor prognosis. For these women it is less clear how many embryos have to be transferred to reduce multiple pregnancies while maintaining optimal pregnancy rates.

In the absence of randomized clinical trials, a more individualized embryo transfer strategy, one that takes the whole prognostic profile of the couple as well as the implantation chances of all available embryos into account, could be a solution. Only two models have been developed that include individual embryo factors while accounting for the correlations in outcome between embryos implanted in the same cycle (15;16). Of these models only one has been externally validated in a single clinical center (15;17). Unfortunately this model was developed for the first IVF cycle only and excluded patients undergoing ICSI. The other prediction model was developed on a larger cohort but was not externally validated and mainly focused on the selection of patients for SET and DET (16). Both models ignore the possibility of combining the aims of preserving or improving pregnancy rates while at the same time lowering multiple pregnancy rates in women with an intermediate prognosis, and especially in women with a poor prognosis and with poor embryo quality, by transferring more embryos.
We developed a multivariable embryo transfer model for cleavage stage embryos on day 3, to calculate the probability of a pregnancy and that of a multiple pregnancy after single, double embryo or triple embryo transfer. The model was built on previously developed and validated models, one based on couple’s characteristics and results from previous cycles, if any, and a second based on day 3 embryo characteristics (18;19).

MATERIAL AND METHODS

Patients
We collected data in a historical cohort of couples that had been treated with IVF or ICSI between January 2004 and April 2011 in the Centre for Reproductive Medicine of the Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands.

All couples in our cohort had been trying to conceive for at least 12 months. They had undergone a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (20). The indication to start IVF or ICSI treatment was determined according to the Dutch IVF guideline (21). If subfertility was caused by tubal pathology, such as two-sided tubal blockage and severe endometriosis, or severe oligozoospermia (post-wash total motile sperm count < 3 million) IVF/ICSI was offered directly (22). In case couples were diagnosed with unexplained subfertility or mild male factor and a good probability of a spontaneous ongoing pregnancy within 12 months, an expectant management for a period of 6 months was maintained (23). In case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia, and unexplained subfertility, at least six intra uterine inseminations (IUI) were applied before IVF/ICSI was offered. In case of ovulation disorders, mainly caused by polycystic ovary syndrome (PCOS), 12 cycles of ovulation induction were applied before IVF/ICSI was offered.

Data on clinical diagnoses, IVF protocol and ovarian response, and laboratory data on embryo morphology and growth, as well as treatment outcomes for all IVF/ICSI cycles were retrieved from our clinical databases and medical records. Included in the analyses were data on stimulated IVF/ICSI cycles and also from frozen–thawed embryo transfers from these stimulated cycles. We excluded IVF/ICSI cycles that involved oocyte or embryo donation, cycles that used surgically retrieved spermatozoa, cycles from human immunodeficiency virus-positive patients, cycles that involved a modified natural cycle and cycles cancelled due to poor ovarian stimulation, ovarian hyperstimulation syndrome or other unexpected medical or non-medical reasons (24). Women underwent controlled ovarian hyperstimulation after down-regulation with the GnRH agonist triptorelin (Decapeptyl®) in a long protocol with a midluteal start. Controlled ovarian hyperstimulation was started on cycle day 5 with recombinant FSH or HMG in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Follicular
maturation was induced by 10,000 IU human chorionic gonadotropin hormone (hCG) (Pregnyl, Organon). Cumulus-oocyte complexes were recovered by transvaginal ultrasound guided follicle aspiration 36 hours thereafter. Oocytes were inseminated with 10,000-15,000 progressively motile spermatozoa (in vitro fertilization) or injected with a single immobilized spermatozoon (intracytoplasmic sperm injection) 2-4 hours after follicle aspiration. Embryos were cultured in Human Tubal Fluid (HTF, Cambrex) or G5 medium (Vitrolife) at 37°C and 5% (HTF) or 6% (G5) CO$_2$ in air (before November 2010) or under 5% O$_2$ (November 2010 until April 2011). Embryo transfer was performed mostly 72h and occasionally 96h after follicle aspiration with a Wallace catheter (Smiths Medical). Supernumerary embryos of good quality were frozen on day 4 after follicle aspiration using a slow-freeze protocol. Luteal phase was supported by progesterone intravaginally two times 200 mg (Utrogestan) per day. A hCG blood test was performed 18 days after oocyte retrieval.

Each embryo was cultured individually. Pronuclear scoring was performed 17-22 hours after insemination/injection and early cleavage was scored 23-28 hours after insemination/injection. On day 2 (41-46 hours after insemination/injection) and day 3 (65-70 hours after insemination/injection) the number of blastomeres was assessed and each embryo was given a morphological score. For the morphological score the degree of fragmentation of the embryo and the uniformity of the blastomeres were assessed (25). The embryo was given a score of 1 (no fragmentation), 2 (<20% fragmentation), 3 (20-50% fragmentation) or 4 (>50% fragmentation). If the blastomeres of the embryo were non-uniform in size, the morphological score was augmented with one point with 4 remaining the lowest possible score. If on day 3 the embryo showed signs of compaction, the embryo was scored as a morula and given a grade based on the degree of compaction (score 1: full compaction, score 2: 50 - <100% compaction and score 3: less than 50% compaction).

**Outcome**

The primary outcome was an ongoing pregnancy. If there was an ongoing pregnancy, we recorded whether it was a single or multiple pregnancy. Ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal ultrasound at a gestational age of at least 10 weeks.

**Data analysis**

We constructed an embryo transfer model to calculate the probability of a pregnancy, a singleton pregnancy and a multiple pregnancy after single, double or triple embryo transfer. We used a so called embryo-uterus model in which survival of the embryos transferred to the uterus depends on their own inherent viability and on the “receptivity” of the uterus that they share (15,26-28).

The embryo transfer model consists of two components: variables specific to the transferred embryo, such as embryo quality and variables specific to the couple, such as
a range of maternal and treatment factors, that affect all embryos equally. For IVF to be successful, both components have to result in a success. For SET, the probability of success is obtained by multiplying the probability of successful implantation of the embryo with the probability of a pregnancy for the couple. With the model, we can also calculate the probability of a twin pregnancy (in case of DET) and that of a twin or a triple pregnancy (in case of TET).

We used a logistic link function for the probabilities of success (15). For our embryo transfer model we used two previously built models with logistic regression modeling: one modeling the chances of success in a next cycle, based on characteristics of the couple and information from previous cycles (IVF model), and a second model to select the best day 3 embryo for transfer (embryo ranking model) (18;19). The latter model only consists of embryo features and calculates a probability of implantation, for each embryo considered for transfer.

The IVF model included the following patient characteristics: female age, duration of subfertility, previous ongoing pregnancy, male subfertility, diminished ovarian reserve, endometriosis, basal FSH, number of failed IVF cycles, an interaction term for female age and male subfertility, and an interaction term for diminished ovarian reserve and endometriosis. For the calculation of pregnancy chances after a failed cycle we added the following embryo features from the previous cycle to the patient characteristics: presence or absence of fertilization, number of embryos, mean morphological score per day 3 embryo, presence of eight-cell embryos on day 3, and presence of morulae on day 3. In the embryo ranking model we included early cleavage, number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morulae on day 3.

For each transfer in the database, we took the linear score for the IVF model, and the linear score for the embryo ranking model, and used a maximum likelihood approach to find two sets of one intercept and a slope each: one intercept and one slope for the IVF score, and one intercept and one slope for the embryo ranking score. The purpose of maximum likelihood is to estimate parameters of the model that best explain the data in the sense of yielding the largest probability (likelihood) of explaining the data.

We also fitted a more complex model, where the intercept and slope were allowed to be different, depending on whether one, two, or three embryos had been transferred. The latter model had eight parameters.

We evaluated the embryo transfer model in terms of discrimination and calibration. Discrimination was expressed as the area under the ROC curve (AUC). Since one, two, or three embryos had been transferred in the cohort used for model development, the outcomes are no ongoing pregnancy, a singleton pregnancy, and a multiple pregnancy. We
Balancing pregnancy and multiple pregnancy chances: an embryo transfer model

Therefore calculated three AUC statistics, corresponding to each of these outcomes. The AUC for no ongoing pregnancy and the AUC for a singleton pregnancy was calculated in all transfers. The AUC for a multiple pregnancy was only calculated for double or triple embryo transfer.

Calibration was evaluated in two ways. The first was by comparing the probability of success (at least one ongoing pregnancy) for all transfers against the observed outcome in five subgroups, defined by calculating quintiles of the calculated probability of success. The second expression of calibration was based on comparing the predicted number of singleton and multiple pregnancies in the SET, DET and TET subgroups against the observed number of pregnancies.

Some of the candidate predictors had missing values. Simple exclusion of couples with missing values on one or more variables commonly causes biased results and decreases statistical efficiency (29). For this reason, missing values in the data were completed by multiple imputation using SPSS (version 18.0). This method uses all available data to impute the missing values based on the correlation between each variable with missing values and all other variables.

Parameter estimation for the embryo transfer was done in R, using the optim function (30). SPSS was used for all other analyses.

RESULTS

In total, 2,201 IVF/ICSI cycles in 1,297 couples could be included in our analyses. In this cohort, 490 single embryo transfers, 1,371 double embryo transfers and 340 triple embryo transfers had been performed, resulting in 415 singletons and 85 multiple pregnancies.

The baseline characteristics are shown in Table I. The average age of the women undergoing IVF/ICSI was 35.5 years with an average duration of subfertility of 3.8 years. Twenty-two percent of the IVF/ICSI cycles had a previous ongoing pregnancy, either after IVF or natural conception.

Embryo transfer model

The slope and intercept for the IVF score were 1.62 and 1.30 respectively; for the embryo ranking score these were 1.55 and 0.90, respectively.

We also fitted a more complex model, where the intercept and slope were allowed to be different, depending on whether one, two, or three embryos had been transferred, but this did not result in a significant improvement in fit (difference in -2LL: 1.966, 4 df, p = 0.74). We therefore decided to use the first, parsimonious model.
Chapter 8

Table I | Baseline characteristics of the cycles

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>n=2201</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age (SD)</td>
<td>35.5 (4.8)</td>
</tr>
<tr>
<td>Duration of subfertility in years (SD)</td>
<td>3.8 (2.5)</td>
</tr>
<tr>
<td>Previous ongoing pregnancy (%)</td>
<td>22%</td>
</tr>
<tr>
<td>FSH (SD)</td>
<td>7.8 (3.6)</td>
</tr>
</tbody>
</table>

Indication for IVF

<table>
<thead>
<tr>
<th>Indication for IVF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexplained subfertility (%)</td>
<td>22%</td>
</tr>
<tr>
<td>Tubal pathology (%)</td>
<td>19%</td>
</tr>
<tr>
<td>Male subfertility (%)</td>
<td>49%</td>
</tr>
<tr>
<td>Polycystic ovarian syndrome (%)</td>
<td>8%</td>
</tr>
<tr>
<td>Diminished ovarian reserve (%)</td>
<td>9%</td>
</tr>
<tr>
<td>Endometriosis (%)</td>
<td>5%</td>
</tr>
<tr>
<td>Cervical hostility (%)</td>
<td>3%</td>
</tr>
</tbody>
</table>

Method of fertilization

<table>
<thead>
<tr>
<th>Method of fertilization</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IVF (%)</td>
<td>47%</td>
</tr>
<tr>
<td>ICSI (%)</td>
<td>53%</td>
</tr>
</tbody>
</table>

| Number of embryos per cycle (SD) | 6.1 (4.1) |
| Number of embryos transferred (SD) | 1.9 (0.6) |
| Number of single embryo transfers | 490 |
| Number of double embryo transfers | 1371 |
| Number of triple embryo transfers | 440 |
| Number of singleton pregnancies | 415 |
| Number of multiple pregnancies | 85 |

Table II shows the recalibrated variables of the embryo transfer model. The most influential variables were: age, presence of diminished ovarian failure or male subfertility, absence or presence of fertilization in the previous cycle, morphological embryo score on day 3 in current cycle, presence of a morula on day 3 in current cycle and previous cycle, number of blastomeres on day 2 in current cycle.

We evaluated the performance of our model by calculation discrimination. The AUCs for no ongoing pregnancy (for all transfers), singleton pregnancy and multiple pregnancy (calculated in the subgroup undergoing DET or TET), were 0.692, 0.665 and 0.778, respectively (Table III).
Balancing pregnancy and multiple pregnancy chances: an embryo transfer model

### Table II | Recalibrated variables of the embryo transfer model

<table>
<thead>
<tr>
<th>Predictors</th>
<th>IVF model $\beta$</th>
<th>Embryo ranking model $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>36.9468</td>
<td>-0.7312</td>
</tr>
</tbody>
</table>

**Patient characteristics**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>IVF model $\beta$</th>
<th>Embryo ranking model $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-3.5478</td>
<td></td>
</tr>
<tr>
<td>Age$^2$</td>
<td>0.1171</td>
<td></td>
</tr>
<tr>
<td>Age$^3$</td>
<td>-0.0013</td>
<td></td>
</tr>
<tr>
<td>Duration of subfertility†</td>
<td>-0.1377</td>
<td></td>
</tr>
<tr>
<td>Previous ongoing pregnancy</td>
<td>0.3215</td>
<td></td>
</tr>
<tr>
<td>Male subfertility</td>
<td>1.4966</td>
<td></td>
</tr>
<tr>
<td>Diminished ovarian reserve</td>
<td>-1.2708</td>
<td></td>
</tr>
<tr>
<td>Endometriosis</td>
<td>-0.7750</td>
<td></td>
</tr>
<tr>
<td>Basal FSH‡</td>
<td>-0.1098</td>
<td></td>
</tr>
<tr>
<td>Number of previous failed IVF cycles</td>
<td>-0.3289</td>
<td></td>
</tr>
<tr>
<td>Age $^*$ male subfertility</td>
<td>-0.0443</td>
<td></td>
</tr>
<tr>
<td>Endometriosis $^*$ diminished ovarian reserve</td>
<td>2.4582</td>
<td></td>
</tr>
</tbody>
</table>

**Embryo parameters previous cycle**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>IVF model $\beta$</th>
<th>Embryo ranking model $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo yes/no after ovum pick up</td>
<td>1.1696</td>
<td></td>
</tr>
<tr>
<td>Number of embryos after ovum pick up$¥$</td>
<td>0.0839</td>
<td></td>
</tr>
<tr>
<td>Mean morphological score all embryos day 3</td>
<td>-0.4970</td>
<td></td>
</tr>
<tr>
<td>Eight-cell embryo yes/no on day 3</td>
<td>-0.4560</td>
<td></td>
</tr>
<tr>
<td>Morulae yes/no on day 3</td>
<td>0.8554</td>
<td></td>
</tr>
</tbody>
</table>

**Embryo parameters this cycle**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>IVF model $\beta$</th>
<th>Embryo ranking model $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early cleavage</td>
<td>0.3854</td>
<td></td>
</tr>
<tr>
<td>Number of blastomeres on day 2 calculated as:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>value of the deviation from 4 $¹$</td>
<td>-0.5140</td>
<td></td>
</tr>
<tr>
<td>Number of blastomeres on day 3 calculated as:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>value of the deviation from 8 $²$</td>
<td>-0.4837</td>
<td></td>
</tr>
<tr>
<td>Morphological score day 3 $³$</td>
<td>-0.8203</td>
<td></td>
</tr>
<tr>
<td>Morula on day 3 $⁴$</td>
<td>-1.8464</td>
<td></td>
</tr>
</tbody>
</table>

$¹$ Number of blastomeres = absolute value (number of blastomeres - 4)

$²$ Number of blastomeres = absolute value (number of blastomeres - 8), morula = 0

$³$ Morula = 0

$⁴$ Presence of morula=1, no morula = 0.

---

[1] Age$^2$ = Age squared

[2] Age$^3$ = Age to the power of 3

[†] duration of subfertility $≥$ 5 years = 5 years

[‡] bFSH $\leq$ 10 IE/L = 10 IE/L

[¥] number of embryos $≥$ 10 = 10 embryos
Table III | Discrimination with the embryo transfer model

<table>
<thead>
<tr>
<th>Outcome</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ongoing pregnancy</td>
<td>0.692</td>
</tr>
<tr>
<td>Single pregnancy</td>
<td>0.665</td>
</tr>
<tr>
<td>Multiple pregnancy</td>
<td>0.778</td>
</tr>
</tbody>
</table>

Calibration is shown in Figure 1 and Table IV. Figure 1 shows that the model is able to distinguish transfers with a low chance of success from those with better chances of success, with slight optimism for high probabilities. Table IV shows the number of singleton and multiple pregnancies in the subgroups of SET transfers, DET transfers and TET transfers separately. The observed and predicted proportion of transfers resulting in singleton and multiple pregnancies were quite similar for the largest group, i.e. DET transfers, and rather comparable for the other two subgroups.

Figure 1 | Calibration plot, showing the association between calculated and observed rates of at least one ongoing pregnancy
Probabilities of success

Our model calculates not only the probability of success of embryo transfer (in terms of establishing an ongoing pregnancy), but also the probability of a singleton or multiple pregnancy with IVF, depending on the number embryos transferred and the specific embryo selected. In Table V we provide a series of hypothetical cases, based on couples and embryos in our database. We assumed that the best embryos (those with the highest scores or probabilities) would be selected for transfer: the best embryo in case of SET, the two best in case of DET, and on the three best for TET.

Table IV | Calibration with the embryo transfer model

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Observed Proportion</th>
<th>Calculated Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SET</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No ongoing pregnancy</td>
<td>0.788</td>
<td>0.820</td>
</tr>
<tr>
<td>Single pregnancy</td>
<td>0.222</td>
<td>0.180</td>
</tr>
<tr>
<td><strong>DET</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No ongoing pregnancy</td>
<td>0.750</td>
<td>0.749</td>
</tr>
<tr>
<td>Single pregnancy</td>
<td>0.195</td>
<td>0.199</td>
</tr>
<tr>
<td>Multiple pregnancy</td>
<td>0.055</td>
<td>0.053</td>
</tr>
<tr>
<td><strong>TET</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No ongoing pregnancy</td>
<td>0.83</td>
<td>0.824</td>
</tr>
<tr>
<td>Single pregnancy</td>
<td>0.157</td>
<td>0.113</td>
</tr>
<tr>
<td>Multiple pregnancy</td>
<td>0.013</td>
<td>0.063</td>
</tr>
<tr>
<td>Patient</td>
<td>Age</td>
<td>Pregnancy history</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>------------------</td>
</tr>
<tr>
<td>A</td>
<td>34</td>
<td>None</td>
</tr>
<tr>
<td>B</td>
<td>42</td>
<td>None</td>
</tr>
<tr>
<td>C</td>
<td>36</td>
<td>Miscarriage not after IVF</td>
</tr>
<tr>
<td>D</td>
<td>27</td>
<td>None</td>
</tr>
</tbody>
</table>
Continuation of table V

<table>
<thead>
<tr>
<th>Embryos from current cycle: three best quality embryos</th>
<th>Pregnancy chances and multiple pregnancy chances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo</td>
<td>Early cleavage</td>
</tr>
<tr>
<td>Patient A</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>Patient B</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>Patient C</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>Patient D</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
</tbody>
</table>

SET = single embryo transfer  
TET = triple embryo transfer  
MPR = multiple pregnancy rates  
DET = double embryo transfer  
PR = pregnancy rates
Chapter 8

**DISCUSSION**

We performed our study to develop a multivariable embryo transfer model for cleavage stage embryos on day 3, to calculate the probability of a pregnancy and that of a multiple pregnancy after single, double embryo or triple embryo transfer. This model was built on two previously developed and validated models in our center, based on couple characteristics, results from previous cycles, if any, and embryo characteristics. In total, 2,201 IVF/ICSI cycles in 1297 couples could be included in our analyses. The model calibrated well after evaluation and had a rather good discriminative capacity.

Several studies developed so-called embryo-uterus models (15-17;26). One model was developed for the first IVF cycle only and excluded ICSI cycles (15;17). This was the only model that was validated externally (15;17). Both models focused on selecting patients for single or double embryo transfer. It could be disputed whether this is the optimal transfer strategy. Women with a poor prognosis and poor embryo quality might require the transfer of more than two embryos to achieve a reasonable probability of a pregnancy (31). As the aim of an optimal transfer strategy should be reducing multiple pregnancies whilst preserving or improving pregnancy rates, only focusing on single and double embryo transfer might lead to substandard care.

Also in 2009 in Europe in more than 17% of the embryo transfers and in the USA in 35% of the embryo transfers three or more embryos were transferred (8;15-17;26;32). By extending our model to triple embryo transfer, and not just restricting it to single- and double embryo transfer, this model may be more clinically useful.

We were also able to include more detailed information on embryo quality from previous IVF cycles and we were able to include more predictive factors on embryo quality for all potential transferrable embryos, this in contrast to the previously developed models (15-17;26).

Our study is not without limitations. The model was based on cohort data, not on data from randomized clinical trials. One of the drawbacks of using cohort data is the possibility of selection bias; data were collected during a period of 7 years and changes in indications for IVF, IVF practice and embryo transfer policies could have affected the influence of predictive factors over time. As there is no data from randomized clinical trials evaluating different embryo transfer strategies for every female age category, data from cohort studies is currently the best evidence available. In our study, the outcome of interest was ongoing (multiple) pregnancy, not (multiple) live birth. Since only 1% to 2% of all ongoing pregnancies result in late miscarriage or still birth we do not expect that our model would fundamentally change when using (multiple) live birth as outcome measure (33).
Another limitation is that we used data of a single center only, and thus the generalizability of the model to other clinics has to be evaluated more extensively in an external validation. We have not yet evaluated whether implementation of this model ultimately leads to a reduction in multiple pregnancies while maintaining or perhaps even improving ongoing pregnancy rates. If the model performs well in external validation current embryo transfer practice should be compared to the use of this model.

Subfertile couples should be well informed about their chances of a pregnancy and multiple pregnancy with different embryo transfer strategies. We feel it is important to note that most couples desire more than one child. When determining the risk/benefit of the number of embryos to be transferred, the number of desired children should also be taken into consideration (34). The time required for multiple frozen cycles in a SET strategy or to initiate a new cycle after a failed IVF/ICSI cycle may result in additional loss of fertility potential. In young women with a good prognosis who have the highest chance of a twin pregnancy, this delay may not be clinically relevant. Yet, for women of older reproductive age this is a relevant concern. For these women a twin pregnancy may be the only chance to conceive and deliver a second child and many studies have shown that a significant proportion of women have a clear preference for a twin pregnancy, regardless of any intervention trying to influence this preference (35-40). So, when deciding how many embryos to transfer, the focus should not be solely on health care costs associated with twin pregnancies, but should equally take into account, after proper counseling on short and long term adverse perinatal outcomes of twin pregnancy, the preference of the couples undergoing treatment. Further research is necessary to define which threshold pregnancy- and multiple pregnancy rates are acceptable, as do methods for discussing this explicitly with patients.

Our prediction model calculates the probability of pregnancy and multiple pregnancy following single-, double- or triple embryo transfer and may aid in the decision how many embryos to transfer, to prevent multiple pregnancies without reducing the chance of pregnancy. Wider applicability of this model requires external validation in databases from other centers.
REFERENCES


Balancing pregnancy and multiple pregnancy chances: an embryo transfer model


Summary and implications for further research
At 11.47 PM on 25 July 1978, the world’s first IVF baby, Louise Joy Brown, was born at Oldham and District Hospital in Greater Manchester. Since the birth of Louise Brown, the number of in vitro fertilization cycles has increased rapidly and it is estimated that today over 5 million IVF babies were born worldwide. IVF is currently one of the most widely used intervention for infertility and the number of cycles is increasing each year even further.

The increase in the number of IVF cycles has not been caused by a sudden epidemic of infertility, but mainly by increased access to IVF and by expansion of the indications for IVF, such as severe male subfertility, unexplained subfertility, cervical hostility, failed ovulation induction, endometriosis, or unilateral tubal pathology. The major difference between the original indication and the ‘newer’ IVF indications is that the couples with bilateral tubal pathology have zero chances of natural conception and completely depend on IVF for getting pregnant, whilst couples with other indications are subfertile, and do have chances of natural conception, which may or may not be better than those with IVF. Due to paucity of data from randomized trials proving the effectiveness of IVF for unexplained infertility, relative to natural conception, the only way to prevent overtreatment at this moment is selection based on the couple’s prognosis. Unfortunately, gynecologists are not able to estimate the probability of achieving a pregnancy with IVF very accurately. Prediction models can help to improve counseling, patient selection and clinical decision making in IVF. Until now, a prediction model that predicts pregnancy chances during the complete IVF procedure, also after failed cycles, and after fresh and frozen-thawed embryo transfer, did not exist.

Soon after the introduction of IVF, pregnancy chances increased but also multiple pregnancy rates. This was caused by the transfer of multiple embryos into the uterine cavity, to increase the likelihood of pregnancy with IVF. The high multiple pregnancy rates caused concern, since multiple pregnancies are associated with an increase in maternal and perinatal morbidity and mortality as well as costs. A logical solution to this recent epidemic of twins would be to transfer only one embryo. Although single embryo transfer (SET) is now an accepted policy for women with a good prognosis, the majority of women currently undergoing IVF have an intermediate or poor prognosis, such as women over 35 with several failed IVF cycles. For these women it is less clear how many embryos have to be transferred to obtain optimal pregnancy rates and low multiple pregnancy rates.

Chapter 1 gives an outline and describes the objectives of this thesis.
Summary and implications for further research

PART ONE: PROGNOSIS WITH IN VITRO FERTILIZATION

In 2009 a systematic review on prediction models in reproductive medicine was published, suggesting that the prediction model developed Templeton et al. in 1996 was the only IVF prediction model with a good predictive performance that could be used for predicting pregnancy chances with IVF. Since the development of this prediction model, IVF has progressed substantially; we thought it questionable whether the model was still valid in current clinical practice.

Chapter 2 describes the external validation of the Templeton prediction model. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic (ROC) curve, and calibration. The area under the ROC curve for the Templeton model was 0.61. Calibration showed a significant and systematic underestimation of success in IVF. Although the Templeton model can distinguish somewhat between women with a high and a low success rate in IVF, it systematically underestimates pregnancy chances and has therefore limited value for current IVF practice.

To develop an improved IVF prediction model the first step we took was the identification of candidate predictors. Chapter 3 systematically reviews nine predictive factors for success in IVF. MEDLINE and EMBASE were systematically searched up to August 2009. Studies were included if they reported an unconditional odds ratio or whenever one could be calculated for one or more of the following factors: age, type of infertility, indication for IVF, duration of infertility, basal FSH, number of oocytes, fertilization method, number of embryos transferred and embryo quality. Fourteen studies were identified. We found negative associations between pregnancy and female age (OR: 0.95, 95% CI: 0.94 to 0.96), duration of subfertility (OR: 0.99, 95% CI: 0.98 to 1.00) and basal FSH (OR: 0.94, 95% CI: 0.88 to 1.00), and a positive association with number of oocytes (OR 1.04, 95% CI: 1.02 to 1.07). No significant association was found for type of infertility and fertilization method. In all studies better embryo quality was associated with higher chances of pregnancy but, as these studies used different factors or combinations of embryo factors to report embryo quality, it was not possible to pool the data and calculate a summary odds ratio. A summary odds ratio could also not be calculated for IVF indication and number of embryos transferred, because studies reporting on these factors used different reference categories. Female age, duration of subfertility, bFSH and number of oocytes, all reflecting ovarian function, are predictors of pregnancy after IVF.

Chapter 4 describes the development of our improved IVF prediction model, which calculates pregnancy chances prior to the start of the first IVF cycle as well as after one or more failed cycles, and takes pregnancy chances after fresh- and frozen-thawed embryo transfer from the same cycle into account. The model was developed on data of 2,621 cycles in 1,326 couples, using multivariable logistic regression and a blockwise
model building strategy. The model was validated in additional data from 440 couples treated between August 2009 and April 2011 at the same center. Thirteen variables were included in the final prediction model: female age, duration of subfertility, previous ongoing pregnancy, male subfertility, diminished ovarian reserve, endometriosis, basal FSH, number of failed IVF cycles, fertilization, number of embryos, mean morphological score per day 3 embryo, presence of eight-cell embryos on day 3, and presence of morulae on day 3. In validation the model had moderate discriminative capacity (c-statistic 0.68, 95% confidence interval: 0.63 to 0.73) but calibrated well, with a range from 0.01 to 0.56 in calculated probabilities. Our prediction model enables the accurate individualized calculation of the probability of an ongoing pregnancy with IVF. This IVF model presented here is able to calculate the chances of an ongoing pregnancy with IVF, both for the first cycle and after any number of failed cycles. Incorporating the model in counselling couples considering IVF may strengthen evidence-based, individualized decision-making and a rational use of scarce resources.

Chapter 5 provides an overview on predictive factors in IVF, the available prediction models in IVF, and provides key principles that can be used to critically appraise the literature on prediction models in IVF. Over 21 papers have reported on the development and/or validation of prediction models in IVF. Of these 21 papers, only two models had a good performance after external validation. Impact analyses have not yet been performed for any of these models. Future research should focus more on updating existing prediction models and adjust or recalibrate them to the local circumstances or setting rather than developing new prediction models.

PART TWO: OPTIMIZING EMBRYO TRANSFER STRATEGIES

In any IVF cycle the question arises which embryo should be transferred, and how many embryos, to obtain optimal pregnancy rates with low multiple pregnancy rates. Chapter 6 presents the results of a tailored embryo transfer policy based on the prognostic profile of the couple. Between August 2006 and April 2011 we adhered to the following embryo transfer protocol at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam: single embryo transfer (SET) was performed followed by double embryo transfer (DET) in frozen embryo transfer cycles in women with a good prognosis (aged <35 years, first cycle, ≥1 top quality embryo). DET was performed in both fresh and frozen cycles in women with an intermediate prognosis (<35 years, first cycle and no top quality embryo available, or aged <35 years and ≥1 failed cycles, or aged 35-38 years). Triple embryo transfer (TET) in both fresh and frozen cycles was performed in women with a poor prognosis (aged ≥39 years). The cumulative ongoing pregnancy rate in women with a good prognosis was 43%, with a multiple pregnancy rate of 2%. In women with an intermediate prognosis this was 27% and 23%, respectively. Corresponding rates were 18% and 13% in women with a poor prognosis. The data in this study can
Summary and implications for further research

be used to guide current practice, by performing SET in women with a good prognosis and TET in women with a poor prognosis. The embryo transfer strategy in women with an intermediate prognosis requires further improvement possibly by refining the transfer strategy according to the ovarian response after ovarian stimulation.

Our final aim was to develop an individualized embryo transfer strategy that takes the complete prognostic profile of the woman into account; to achieve optimal pregnancy rates at low risk of multiple pregnancies. The first step in the development of such a model was to evaluate embryo quality more extensively. Chapter 7 reports on the development of a prediction model to rank embryos within a single IVF/ICSI cycle according to their ongoing implantation potential. We prospectively studied embryo transfers on day 3 between January 2004 and July 2009 at the Center for Reproductive Medicine at the Academic Medical Center, University of Amsterdam. We evaluated pronuclear score, early cleavage, number of blastomeres on day 2 and day 3, morphological score on day 2 and day 3 as potential predictors for implantation. A model was developed using multivariable logistic regression. The prediction model was externally validated on embryo transfer data collected between August 2009 and September 2011 in the same center. Five factors were included in the final prediction model: early cleavage, number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morulae on day 3. In the external validation the model showed moderate discriminative capacity (c-statistic 0.70) and calibrated well. The model was able to distinguish embryos with high ongoing implantation potential from embryos with moderate or low ongoing implantation potential. The model can be used by embryologists as an objective tool to rank embryos according to their implantation potential thereby aiding the selection of embryos for transfer.

Chapter 8 reports on the development of a multivariable embryo transfer model for cleavage stage embryos on day 3 to calculate the probability of a pregnancy and the chances of multiple pregnancy after single, double embryo or triple embryo transfer. The embryo transfer model consists of two components: variables specific to the transferred embryo(s), such as embryo quality and variables specific to the couple, such as a range of maternal factors, treatment and cycle data that affect all embryos equally. We combined two previously described models, one for ranking embryos and one for estimating pregnancy chances. The performance of the prediction model was assessed in terms of discrimination, i.e. the area under the receiver operation characteristic curve (AUC) and calibration.

In total, 2,201 IVF/ICSI cycles in 1,297 couples could be included in our analyses. The embryo transfer model includes the following variables specific for the couple: female age, duration of subfertility, previous ongoing pregnancy, male subfertility diminished ovarian reserve, endometriosis, basal FSH and number of failed IVF cycles and in case of a failed IVF cycle: presence of absence fertilization, number of embryos, mean morphological score,
presence of 8-cell embryos and presence of morulae on Day 3 of the previous cycle, and the following variables specific for the actual IVF cycle: early cleavage, number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morulae on day 3. The model calibrated well. The AUC for predicting pregnancy was 0.692 and multiple pregnancy was relatively high, at 0.778.

Our model may aid in the decision how many embryos to transfer in any specific woman with any specific embryos to prevent a multiple pregnancy without reducing the chances of a singleton pregnancy. Wider applicability of this model requires external validation in databases from other centers.

IMPLICATIONS OF THIS THESIS

The prediction models discussed in this thesis could help to improve every day IVF practice. Before the introduction of prediction models, most clinical decisions and embryo selection procedures were based on expert opinion. With the prediction models presented in this thesis a more evidence-based treatment strategy is possible, and also a more ‘patient tailored’ treatment. Imagine a couple with an indication for IVF/ICSI because of unexplained subfertility. Their fertility specialist will calculate, with the IVF model (Chapter 4), their chances of a pregnancy with IVF/ICSI. Together with their fertility specialist, the couple can decide if their chances outweigh the burden and risk of an IVF cycle. If they decide to start with IVF, have undergone oocyte retrieval, and embryos have developed, the embryologist can rank the embryos by their implantation potential, with the embryo model (Chapter 7). Before the embryo transfer, the embryologist can calculate, with the embryo transfer model (Chapter 8), their pregnancy chances and multiple pregnancy chances, with the best embryo or with different combinations of embryos. If they fail to get pregnant, the fertility specialist will calculate again, with the IVF model (Chapter 4), their chances of a pregnancy in the subsequent cycle.

We suggest that IVF prediction models can potentially lead to more efficient practice, prevent overtreatment, improve embryo selection and lower multiple pregnancy rates. Yet, three important questions remain: how high should the probability of a pregnancy with IVF be before starting or continuing treatment, what are acceptable multiple pregnancy rates, and do these rates differ between women with different prognostic profiles?

The threshold for starting or continuing an IVF cycle is not described in any guideline and is definitely not set in stone. It is very likely that this threshold will differ between different stakeholders, between subfertile couples, fertility specialists or insurance companies. Further research is necessary to explore this threshold and factors driving the differences.
Summary and implications for further research

It is also unclear what an acceptable multiple pregnancy rate is. When considering the risks and benefits associated with the number of embryos to be transferred, the number of desired children should also be taken into consideration (1). Most couples aspire to have more than one child. As fertility potential decreases with increasing female age, older women with an intermediate or poor prognosis often do not have enough time for a second child. For these women a twin pregnancy may be the only chance to conceive and deliver two children. Many studies have shown that a significant proportion of women have a clear preference for a twin pregnancy, regardless of any intervention trying to influence this preference (2-7). So, when deciding how many embryos to transfer, the focus should not be solely on associated health care costs with twin pregnancies, but should equally take into account the preference of the couples undergoing treatment.

The development of prediction models is a continuous process; as patient populations may shift during the years the group of patients used for the development and validation of prediction model may differ from the current population. Reproductive techniques may also evolve and new biomarkers with predictive value may become available. To safeguard valid predictions for future patients, regularly updating and adapting the prediction model to the new setting is necessary. The best way to achieve this is may be by developing national electronic databases to collect large numbers of IVF cycles, so that models can be updated more easily.

Until larger randomized controlled trials comparing IVF with natural conception and different embryo transfers strategies in all prognostic categories are performed, we believe that incorporating prediction models in counselling can strengthen the evidence-based, individualized decision-making and lead to a more rational use of scarce resources.
REFERENCES


Samenvatting en implicaties voor toekomstig onderzoek

De toename van het aantal IVF-cycli is niet veroorzaakt door een plotselinge onvruchtbaarheidsepideem, maar doordat IVF beter toegankelijk is geworden voor subfertiele paren en doordat IVF ook is ingezet voor andere indicaties dan dubbelzijdige tubopathologie, zoals ernstige mannelijke subfertiliteit, onverklaarde subfertiliteit, cervix factor, mislukte ovulatie-inductie, endometriose of eenzijdige tubopathologie. Het grote verschil tussen de oorspronkelijke IVF indicatie en de nieuwere IVF indicaties is dat paren met dubbelzijdige tubopathologie niet of nauwelijks kans hebben om op natuurlijke wijze zwanger te worden (infertiele paren) en daarmee volledig van IVF afhankelijk zijn, terwijl paren met nieuwere IVF indicaties subfertiel zijn en dus nog wel een spontane zwangerschapskans hebben. Deze spontane kans op een zwangerschap kan dan beter of slechter zijn dan de kans op zwangerschap met IVF.

Er zijn nauwelijks gegevens van gerandomiseerde studies die de effectiviteit van IVF bij onverklaarde onvruchtbaarheid aantonen. De enige manier om overbehandeling in deze paren te voorkomen is ze al dan niet te behandelen met IVF op basis hun prognose.

Helaas blijken artsen niet goed in staat te zijn de kansen op een spontane zwangerschap of een zwangerschap na IVF te schatten. Predictiemodellen kunnen als hulpmiddel dienen voor het berekenen van deze zwangerschapskansen. Tot nu toe bestaat er echter geen predictiemodel dat de kans op zwangerschap voorspelt tijdens de volledige IVF-procedure, dat wil zeggen de zwangerschapskans na elke mislukte IVF cyclus en na terugplaatsing van verse of ingevroren en later ontdooide embryo’s.

Al snel na de invoering van IVF namen niet alleen de zwangerschapskansen per behandeling toe, maar ook het aantal meerlingzwangerschappen. Dit werd veroorzaakt doordat bij IVF vaak verschillende embryo’s in de baarmoeder worden geplaatst om zo de kans op zwangerschap te verhogen. De zwangerschapskansen stegen hiermee, maar de keerzijde van de medaille was dat dit gepaard ging met hogere kansen op meerlingzwangerschappen. Het hoge aantal meerlingzwangerschappen baarde ernstig zorgen aangezien meerlingzwangerschappen gepaard gaan met hogere maternale en perinatale morbiditeit en mortaliteit, evenals hogere gezondheidszorgkosten. Een logische oplossing voor deze meerlingeproblematiek zou de terugplaatsing van één embryo zijn. Momenteel wordt terugplaatsing van één embryo als een valide beleid beschouwd voor vrouwen met een goede prognose. De meerderheid van de vrouwen die IVF ondergaan hebben echter een intermediaire of een slechte prognose. Dat zijn vrouwen die ouder zijn
Samenvatting en implicaties voor toekomstig onderzoek

dan 35 en/of vrouwen met meerdere mislukte IVF cycli. Voor deze vrouwen is het minder
duidelijk hoeveel embryo’s teruggeplaatst moeten worden om een kans op zwangerschap te
krijgen met een zo laag mogelijke kans om op meerlingzwangerschap.

**Hoofdstuk 1** geeft een beschrijving van de achtergrond en doelstellingen van dit
proefschrift.

**DEEL EEN: PROGNOSE MET IN VITRO FERTILISATIE**

In 2009 werd een systematische overzichtsartikel over predictiemodellen in de
voortplantingsgeneeskunde gepubliceerd. Uit dit artikel bleek dat er slechts een
predictiemodel was dat in de klinische praktijk gebruikt kon worden voor het voorspellen
van zwangerschapskansen met IVF. Dit model was in 1996 ontwikkeld door Templeton
et al. Sinds de ontwikkeling van dit model zijn de zwangerschapskansen met IVF en de
technieken voor IVF aanzienlijk verbeterd. Wij vroegen ons dan ook af of het model
ontwikkeld door Templeton et al. nog wel bruikbaar was is de huidige klinische praktijk.

**Hoofdstuk 2** beschrijft de externe validatie van het Templeton predictiemodel. De kwaliteit
van het predictiemodel werden geëvalueerd door het onderscheidend vermogen van het
model te analyseren op basis van de oppervlakte onder de ‘receiver operator characteristic’
(ROC) curve en door calibratie. De oppervlakte onder de ROC curve was 0,61. De
calibratiecurve toonde een systematische onderschatting van de kans op succes met IVF.
Hoewel het Templeton model enigszins onderscheid kan maken tussen paren met een hoge
een die met een lage kans op succes met IVF, onderschat het model systematisch de kansen
en heeft het dus geen echte toegevoegde waarde meer voor de huidige IVF-praktijk.

De eerste stap die wij namen om een verbeterd en vernieuwd IVF predictiemodel te
ontwikkelen was het evalueren van potentiele voorspellende factororen.

In **hoofdstuk 3** laten we een systematisch literatuuroverzicht zien van negen voorspellende
factoren voor succes bij IVF. MEDLINE en EMBASE werden doorzocht tot en met
augustus 2009. Onderzoeken kwamen in aanmerking als ze een univariabele odds ratio
rapporteerde of als deze berekend kon worden voor de volgende factoren: leeftijd, primaire
of secundaire subfertiliteit, indicatie voor IVF, duur van de subfertiliteit, basaal FSH, aantal
oocyten, methode van fertilisatie, aantal teruggeplaatste embryo’s en embryo kwaliteit.
Veertien studies werden geïncludeerd. We vonden significante negatieve associaties tussen
de kans op zwangerschap en de leeftijd van de vrouw, duur van de subfertiliteit en basale
FSH en een positieve associatie met het aantal oocyten. Er werd geen associatie tussen
zwangerschap en primaire/secundaire subfertiliteit, methode van fertilisatie gevonden.
Betere embryokwaliteit was geassocieerd met een hogere kans op zwangerschap, er
kon echter geen meta-analyse verricht worden omdat de onderzoeken verschillende
embryofactoren of combinatie van embryofactoren gebruikten om embryokwaliteit te
rapporteren. Er kon ook geen meta-analyse verricht worden voor indicaties voor IVF en het aantal embryo’s. Samenvattend bleken de leeftijd van de vrouw, de duur van de subfertiliteit, bFSH en het aantal oocyten voorspellers te zijn van de kans op zwangerschap met IVF.


In het definitieve multivariaabele model werden 13 factoren opgenomen: leeftijd van de vrouw, duur van de subfertiliteit, eerdere doorgaande zwangerschappen, mannelijke subfertiliteit, verminderde ovariële reserve, endometriose, basaal FSH, het aantal mislukte IVF cycli, fertilisatie, aantal embryo’s in de vorige IVF cyclus, de gemiddelde morfologische score per dag 3 embryo in de vorige IVF cyclus, de aanwezigheid van achtcellige embryo’s en van morulae op dag 3 in de vorige IVF cyclus. In de externe validatie had het model een matig discriminerend vermogen (c-statistiek 0,68, 95% BI: 0,63-0,73) en calibreerde het model goed. De voorspelde kansen hadden een spreiding van 0,01-0,56.

Ons IVF-model kan nauwkeurig de kans op een doorgaande zwangerschap met IVF voorspellen, na verse- en ingevroren-ontdooide embryo transfer, voor de eerste cyclus en na een aantal mislukte cycli. Het model kan als hulpmiddel dienen in de spreekkamer om paren te adviseren bij de beslissing om al dan niet te starten of door te gaan met IVF. Tevens kan het dienen als leidraad voor de behandeldend arts bij het maken van een behandelplan.

Hoofdstuk 5 geeft een overzicht van voorspellende factoren voor het resultaat na IVF, de beschikbare predictiemodellen voor IVF en biedt belangrijke basisprincipes die gebruikt kunnen worden om de literatuur over predictiemodellen voor IVF kritisch te beoordelen. Meer dan 21 artikelen rapporteerden over de ontwikkeling en / of validatie van predictiemodellen voor IVF. Deze 21 artikelen beschreven slechts twee kwalitatief goede modellen, na externe validatie. Er zijn nog geen impactanalyses uitgevoerd voor deze modellen. In plaats van het ontwikkelen van nieuwe predictiemodellen zou toekomstig onderzoek zich meer moeten richten op het ‘up to date’ houden van bestaande predictiemodellen en het aanpassen of opnieuw calibreren van deze modellen, zodat ze toepasbaar worden voor de plaatselijke omstandigheden of instelling.
DEEL TWEE: HET OPTIMALISEREN VAN EMBRYOTRANSFERSTRATEGIEËN

Bij elke IVF-cyclus rijzen steeds twee vragen: welke en hoeveel embryo’s moeten worden terug geplaatst voor optimale zwangerschapskansen met lage kansen op meerlingzwangerschappen. **Hoofdstuk 6** laat de resultaten zien van een op maat gemaakte embryotransferbeleid gebaseerd op het prognostische profiel van het paar. Tussen augustus 2006 en april 2011 hebben we het volgende embryotransferprotocol gebruikt bij het Centrum voor Voortplantingsgeneeskunde in het Academisch Medisch Centrum, Universiteit van Amsterdam. Bij vrouwen met een goede prognose (leeftijd <35 jaar, eerste cyclus, ≥ 1 top kwaliteit embryo) werd één embryo (SET) teruggeplaatst, indien niet succesvol werden er twee embryo’s (DET) teruggeplaatst bij ingevroren-ontdooide embryo cycli. DET werd uitgevoerd bij zowel verse als ingevroren-ontdooide cycli bij vrouwen met een intermediaire prognose (<35 jaar, eerste cyclus en geen top kwaliteit embryo beschikbaar, of de leeftijd <35 jaar en ≥ 1 mislukte cycli, of 35-38 jaar oud). Drie embryo’s (TET) werden terug geplaatst in zowel de verse als ingevroren-ontdooide cycli bij vrouwen met een slechte prognose (leeftijd ≥ 39 jaar). Het cumulatieve percentage doorgaande zwangerschappen bij vrouwen met een goede prognose was 43%, met een meerling zwangerschapspercentage van 2%. Bij vrouwen met een intermediaire prognose was dit respectievelijk 27% en 23%. Bij vrouwen met een slechte prognose was dit 18% en 13%. De gegevens in deze studie kunnen als leidraad gebruikt worden, door een embryo terug te plaatsen bij vrouwen met een goede prognose en drie embryo’s terug te plaatsen bij vrouwen met een slechte prognose. De embryotransferstrategie bij vrouwen met een intermediaire prognose vereist verdere verfijning, mogelijk op basis van bijvoorbeeld ovariële respons na ovariële hyperstimulatie.

Ons uiteindelijke doel was om een geïndividualiseerde embryoterugplaatsstrategie te ontwikkelen die rekening houdt met het volledige prognostische profiel van het paar, om zo optimale zwangerschapskansen te bewerkstellingen met lage kansen op een meerlingzwangerschappen. De eerste stap in de ontwikkeling van een dergelijk model was het uitgebreider evalueren van de embryokwaliteit.

**Hoofdstuk 7** beschrijft de ontwikkeling van een predictiemodel voor het rangschikken van embryo’s binnen een IVF / ICSI op basis van hun kans op implantiatie. In een prospectief onderzoek evalueerden wij de embryoterugplaatsingen op dag 3 tussen januari 2004 en juli 2009 in het Centrum voor Voortplantingsgeneeskunde in het Academisch Medisch Centrum, Universiteit van Amsterdam. We evalueerden de volgende potentiële voorspellende factoren voor implantiatie: pronuclei score, vroege celdeling, aantal blastomeren op dag 2 en dag 3, morfologische score op dag 2 en dag 3. Het model werd ontwikkeld met behulp van multivariabele logistische regressie. Het predictiemodel werd extern gevalidateerd op de gegevens van de embryoterugplaatsingen tussen augustus 2009 en september 2011 verzameld in hetzelfde centrum. Vijf factoren werden opgenomen in het definitieve model: vroege celdeling, aantal blastomeren op dag 2, aantal blastomeren
Hoofdstuk 8 beschrijft de ontwikkeling van een embryoterugplaatsmodel op dag 3 dat de zwangerschapskans en de kans op meerlingzwangerschap berekent na de terugplaatsing van een, twee of drie embryo's. Het model bestaat uit twee componenten: variabelen die specifiek zijn voor de terug geplaatste embryo's, zoals de kwaliteit van het embryo, en de variabelen die specifiek zijn voor het paar, zoals een reeks van maternale factoren, behandeling en cyclusgegevens die van invloed zijn op alle embryo's. We combineerden de twee eerder beschreven modellen: een voor het rangschikken van embryo's op basis van de implantatiekansen en een voor het schatten van de zwangerschapskansen met IVF. De kwaliteit van het model werd beoordeeld door middel van discriminatie, dat wil zeggen de oppervlakte onder de 'receiver operation characteristic curve' (AUC) en calibratie.

In totaal konden 2201 IVF/ICSI-cycli van 1297 paren worden gebruikt voor onze analyses. Het embryoterugplaatsmodel bestond uit de volgende variabelen specifiek voor het paar: leeftijd van de vrouw, duur van de subfertilititeit, eerder doorgaande zwangerschappen, mannelijke subfertilititeit, vermindere ovariële reserve, endometriose, basaal FSH en het aantal mislukte IVF-cycli en, in het geval van een mislukte IVF-cyclus: aanwezigheid van afwezigheid fertilisatie, aantal embryo's, gemiddelde morfologische score, aanwezigheid van 8-cellige embryo's en de aanwezigheid van morulae op dag 3 van de vorige cyclus. Daarbij werden opgenomen de volgende variabelen die specifiek zijn voor de huidige IVF-cyclus: early cleavage, aantal blastomeren op dag 2, aantal blastomeren op dag 3, de morfologische score op dag 3 en de aanwezigheid van morula op dag 3. Het model calibreerde goed. De AUC voor het voorspellen van de kans op zwangerschap was 0,692 en voor de kans op een meerlingzwangerschap was dit relatief hoog: 0,778. Ons model kan als hulpmiddel gebruikt worden bij de beslissing hoeveel embryo's teruggeplaatst moeten worden in paren om zo de een meerlingzwangerschap te voorkomen, zonder dat de kans op een eenling zwangerschap te verminderen. Bredere toepasbaarheid van dit model vereist externe validatie op gegevens van andere centra.
Samenvatting en implicaties voor toekomstig onderzoek

**IMPLICATIES VAN DIT PROEFSCHRIFT**

De besproken predictiemodellen in dit proefschrift kunnen helpen om de dagelijkse IVF-praktijk te verbeteren. Voordat de invoering van deze modellen werden de meeste klinische beslissingen en embryo selectie procedures gebaseerd op adviezen van experts. De predictiemodellen gepresenteerd in dit proefschrift maken een meer rationele behandeling strategie mogelijk, die tevens meer is toegesneden op de individuele patiënt. Stelt u zich een paar voor en een indicatie voor IVF/ICSI op basis van onverklaarde subfertilititeit. De behandendel arts zal met behulp van het IVF-model (hoofdstuk 4), kunstwassen op een zwangerschap met IVF/ICSI berekenen. Samen met de arts kan het paar beslissen of deze kansen op zwangerschap met IVF opwegen tegen de lasten en risico’s van een IVF-cyclus. Als ze dan besluiten om te starten met IVF en er een ecelpunctie heeft plaats gevonden met het ontstaan van embryo’s, kan de embryoloog de embryo’s rangschikken op basis van hun individuele implantatiekans middels het embryo-model (hoofdstuk 7). Voordat de embryoterugplaatsing plaats vindt kan de embryoloog de kans op zwangerschap berekenen voor verschillende combinaties van embryo’s met behulp van het embryoterugplaatsmodel (hoofdstuk 8). Als er geen zwangerschap ontstaat, kan de arts opnieuw hun zwangerschapskans bij een eventuele volgende IVF-cyclus berekenen met behulp van het IVF-model (hoofdstuk 4).

Wij denken dat IVF predictiemodellen leiden tot een effectievere en efficiëntere behandeling, zodat overbehandeling voorkomen kan worden, er betere embryo selectie plaats vindt en er minder meerlingzwangerschappen zullen ontstaan. Drie belangrijke vragen blijven echter bestaan: hoe hoog moet de zwangerschapskans zijn voor het starten of voortzetten van behandeling, wat vinden wij een aanvaardbaar risico voor een meerlingzwangerschap, en verschilt de hoogte van dit risico tussen vrouwen met verschillende prognostische profielen?

Er is nog geen consensus over de hoogte van de drempel voor het starten of voortzetten van een IVF-cyclus. Het is zeer waarschijnlijk dat deze drempel zal verschillen tussen de verschillende belanghebbenden: subfertiele paren, artsen of verzekeringsschappijen. Verder onderzoek is nodig om te onderzoeken hoe hoog deze drempel moet zijn en welke factoren de hoogte van de drempel beïnvloeden.

Het is ook nog onduidelijk wat een acceptabel risico is op een meerlingzwangerschap. Bij deze overweging is het belangrijk om het aantal gewenste kinderen door een paar mee te laten wegen (1). De meeste paren streven naar meer dan een kind. De vruchtbaarheid neemt af met de stijgende leeftijd van de vrouw; oudere vrouwen met een intermediaire of slechte prognose hebben dus vaak niet genoeg tijd voor een tweede kind. Voor deze vrouwen kan een tweelingzwangerschap de enige kans zijn op twee kinderen. Vele studies hebben aangetoond dat een aanzienlijk deel van de vrouwen een duidelijke voorkeur hebben voor een tweelingzwangerschap, deze wens persisteert zelfs nadat meerdere
interventies uitgeprobeerd zijn om deze voorkeur te beïnvloeden (2-7). Wanneer er besloten moet worden over het aantal embryo’s dat teruggeplaatst wordt, moet de nadruk niet uitsluitend gebaseerd worden op potentiële kosten van de gezondheidszorg geassocieerd met tweelingzwangerschappen, maar moet evenzeer rekening worden gehouden met de voorkeur van de paren die een behandeling ondergaan.

De ontwikkeling van predictiemodellen is een continu proces; de patiëntenpopulatie kan veranderen in de loop der jaren zodat deze kan afwijken van de patiëntpopulatie die gebruikt is voor de ontwikkeling en validatie van het predictiemodel. Voortplantingstechnieken kunnen zich ook verder ontwikkelen en nieuwe biomarkers met voorspellende waarde kunnen beschikbaar komen. Om valide voorspellingen voor toekomstige patiënten te kunnen waarborgen, moeten de modellen regelmatig worden bijgewerkt en aangepast worden. De beste manier om dit te bereiken is het ontwikkelen van een nationale elektronische databank om grote aantallen IVF-cycli te verzamelen, zodat modellen gemakkelijker kunnen worden bijgewerkt.

Totdat er grotere gerandomiseerde studies worden uitgevoerd die IVF vergelijken met afzien van IVF en verschillende embryoterugplaatsstrategieën in alle prognostische categorieën, kan de integratie van predictiemodellen in de advisering van paren met een onvervulde kinderwens, geïndividualiseerde besluitvorming versterken en leiden tot een rationeler gebruik van schaarse middelen.
Appendices
Portfolio
PHD TRAINING

General courses
2007  Evidence based searching (0.1 ECTS)
2009  Scientific Writing in English for Publication (1.5 ECTS)
2011  BROK (‘Basiscursus Regelgeving Klinisch Onderzoek’) (0.9 ECTS)

Specific courses
2008  Practical Biostatistics (1.1 ECTS)
2009  NIHES, Prognostic research (1.5 ECTS)
       NIHES, Advanced Analysis of Prognosis Studies (0.9 ECTS)
2010  Advanced Topics in Biostatistics (1.1 ECTS)
       Clinical Epidemiology (0.6 ECTS)

Seminars
2007-2012  Weekly department seminars (7.6 ECTS)

Presentations
2005  Poster presentation at the Dutch Atherosclerosis Society (0.5 ECTS)
2009  Poster presentation at the American Society for Reproductive Medicine (ASRM), Atlanta (0.5 ECTS)
2010  Oral presentation at the European Society of Human Reproduction and Embryology (ESHRE), Rome (0.5 ECTS)
2011  Poster presentation at the American Society for Reproductive Medicine (ASRM), Orlando (0.5 ECTS)
2011  Oral presentation at the refereeravond obstetrics and gynaecology cluster Amsterdam (0.5 ECTS)
2012  Oral and poster presentation at the European Society of Human Reproduction and Embryology (ESHRE), Istanbul (1.0 ECTS)

(International) national conferences
2008  Reinier de Graaf Symposium  Embryo Quality, Groningen (0.5 ECTS)
2009  European Society of Human Reproduction and Embryology (ESHRE), Amsterdam (0.75 ECTS)
2010  European Society of Human Reproduction and Embryology (ESHRE), Rome (1.0 ECTS)
2012  European Society of Human Reproduction and Embryology (ESHRE), Istanbul (0.75 ECTS)

Other
2007-2012  Journal Club (3.8 ECTS)
TEACHING
Tutoring/mentoring
Student coaching/mentoring scientific research project 2008 (1.0 ECTS)
Student coaching/mentoring scientific research project 2009 (1.0 ECTS)

PARAMETERS OF ESTEEM
Grants
2010 NutsOhra €25,000

LIST OF PUBLICATIONS


Portfolio

L.L. van Loendersloot, S. Repping, P.M.M. Bossuyt, F. van der Veen, M. van Wely. Prediction models in in vitro fertilization; where are we? *Journal of Advanced Research* 2013; *in press: May 9.*


Dankwoord
Dankwoord

Het boekje is dan eindelijk klaar, wat een heerlijk gevoel! Dit proefschrift was er niet gekomen zonder jullie steun. Promoveren doe je immers niet alleen. Een aantal mensen wil ik in het bijzonder bedanken.

Beste Professor Van der Veen, beste Fulco, wat heb ik in de afgelopen jaren toch ontzettend veel van je geleerd. Jouw bezieling voor wetenschappelijk onderzoek is groot en daardoor zeer motiverend. Keer op keer verbaas je me met je heldere visie en met het voor ogen houden van de ‘big picture’. Dit alles met als belangrijkste doel: het verbeteren van de patiëntenzorg. Je leermethodes zijn soms wat ONORTHODOX, echter het dient zijn doel. Ik voel me bevoorrecht dat ik onder jouw vleugels onderzoek heb kunnen doen. Je hebt in belangrijke mate bijgedragen aan mijn wetenschappelijke ontwikkeling en mijn ontwikkeling als arts. Heel veel dank hiervoor!

Beste Professor Bossuyt, beste Patrick, ook wel het orakel van het AMC. Toen ik begon met dit proefschrift was ik zo groen als gras op het gebied van predictiemodellen. Ik was immers maar een simpele arts met de statistische kennis van een middelbare scholier. Het was in het begin toch wel erg lastige materie en na ons overleg moest ik regelmatig even bij Madelon of een andere collega langs om te checken of ik überhaupt wel had begrepen wat we besproken hadden. Door jouw rust en goede uitleg heb ik mijn kennis op dit gebied flink kunnen uitbreiden. Ik wil je bedanken voor al je tijd, geduld en met name ook de kennis die je mij hebt bijgebracht. Ik weet zeker dat ik hierdoor niet alleen een betere onderzoeker ben geworden maar ook een betere arts!

Beste Professor Repping, beste Sjoerd, waar Fulco de Engelse stijl nastreeft, ben jij meer van de Amerikaanse stijl. Een manuscript mag best glossy en gepimpt zijn. Moeilijke problemen worden door jou begrijpelijk gemaakt door middel van een simpel diagrammetje of eenvoudig tekeningentje. Wat mij betreft ligt daar een groot deel van je kracht. Je bent een echte onderzoeker in hart en nieren en het is heerlijk om met jou te sparen over onderzoek; het liefst met een biertje in de hand leunend tegen de bar. Dank voor al je tijd en met name database uitwisselingen Sjoerd…….

Lieve dr. Van Wely, lieve Madelon, tussen alle alfamannetjes was het heerlijk om er een relaxte vrouw bij te hebben, die zich ogenschijnlijk nergens druk om maakt. Als ik even door de bomen het bos niet meer zag dan was jij er altijd om me net dat zetje te geven om door te gaan. Ook was het fijn om het met je over niet werk gerelateerde zaken te spreken. Ik waardeer je inzet en je positieve energie enorm. Dank hiervoor!

De leden van de promotiecommissie: Dr. Dumoulin, Prof. dr. Lambalk, Prof. dr. Mol en Prof. dr. Steyerberg wil ik bedanken voor het beoordelen van het manuscript en het plaatsnemen in de commissie. I would like to thank Prof. dr Bhattacharya for crossing the North Sea and for his presence here today. As one of the leading experts in the field of IVF it is a great honour that you are one of my opponents.
Lieve Anna, wat heb ik het toch met jou getroffen tijdens mijn onderzoeks- en CVV-tijd! Ik kijk regelmatig met veel plezier terug naar onze tijd samen. Ik heb zoveel met jou en ook om jou kunnen lachen dat ik er nu, terwijl ik dit schrijf, alweer tranen van in mijn ogen krijg. Jouw positieve levensinstelling, je open en warme karakter maken je echt een geweldig mens. Ik ben blij en trots om jou als vriendin te hebben en het is dan ook een eer dat je vandaag als paranimf naast me staat!

Lieve Suthesh, wat ben ik toch blij dat we ooit onze wetenschappelijke stage op de vasculaire geneeskunde samen zijn gaan doen. Vanaf dat moment heeft onze vriendschap echt een vlucht genomen. Met jouw nuchtere gevoel voor humor help je mij zaken in perspectief te plaatsen. Je magische woorden voor mijn eerste echte presentatie op een groot internationaal congres staan in mijn geheugen gegrift: “Ach Laura, het aller ergste dat kan gebeuren is dat men zegt: slecht verhaal maar een leuk koppie”. Dat jij ook een goede en dierbare vriend bent in minder gezellige tijden heb je het afgelopen jaar ook zeker bewezen. Suthesh je bent een topper. Wat fijn dat jij vandaag naast me staat!

Lieve meisjes van de CVV, Alexandra, Emmy en Noortje, wat een feest was het met jullie op de CVV te zitten en onderzoek met jullie te kunnen doen. Het is voor mij buitengewoon fijn geweest dat ik met jullie zowel hoogte als diepte punten heb kunnen delen. Jullie zijn de meest attente mensen die ik ken, in goede maar zeker ook in mindere tijden. Zoals ik al schreef promoveren doe je nooit allen, maar zonder leuke collega’s is het gewoon niet te doen. Dank voor al jullie gezelligheid waar hopelijk nog veel van gaat komen…..

Lieve kamergenootjes van H4-240, Elisabeth & Stef. Onze onderzoekskamer of misschien beter gezegd ons ‘hok’ was toch echt DE gezelligste kamer van het AMC! Het was fijn om onderzoekers om me heen te hebben die wat verder in het traject zaten en mij een beetje weg wijs konden maken. Elies, je bent een hele lieve en goedlachse collega. Stef, wat was het een feest om een man met veel humor in ons kippenhok te hebben.

Alle onderzoekers van het AMC (jullie zijn tegenwoordig met teveel om bij naam te noemen…) wil ik bedanken voor alle gezelligheid en tips & tricks! Dankzij jullie heb ik mijn onderzoek met veel plezier kunnen doen.

Lieve collega’s IVF-artsen, verpleegkundigen en staf van de CVV, door jullie heb ik altijd met heel veel plezier op de CVV gewerkt. Voor mijn gevoel blijft de CVV toch een soort ‘kleine familie’. Stiekem mis ik af en toch nog onze veel te kleine koffiekamer op A1. Ik heb me door jullie altijd enorm welkom en gewaardeerd gevoeld. Dank voor de leuke tijd als IVF-arts!

Lieve Nino, of moet ik zeggen ‘doctor Feel Good’! Jij bent het hart van de CVV. Je hebt me ontzettend veel geleerd en ik heb echt enorm met je gelachen. Gelukkig heb ik de eer om nogmaals in mijn opleiding met jou te werken tijdens mijn stage op de CVV. Daar kijk ik nu al naar uit!!!
Dankwoord

Lieve collega’s van het Sint Lucas Andreas ziekenhuis (verpleegkundigen, verloskundigen, doktersassistenten, secretaresses, arts-assistenten en gynaecologen) vanaf mijn eerste werkdag voel ik mijn enorm welkom en thuis in het SLAZ. Het SLAZ is door jullie een hele fijne en leuke werkplek en mede hierdoor heb ik de rust gehad om de laatste hand aan mijn proefschrift te leggen. Ik ben echt een geluksvogel dat ik mijn opleiding heb kunnen starten in het SLAZ.

Er zijn veel mensen die niet betrokken zijn geweest bij mijn onderzoek, maar wel heel belangrijk zijn in mijn leven. Al mijn lieve vrienden en vriendinnen, ook jullie zijn er altijd voor mij; tijdens leuke tijden en belangrijker nog, ook tijdens minder leuke tijden. Of we samen lachen, feesten, eten of goede gesprekken voeren, jullie zorgen ervoor dat het leven leuk is. Wat zou ik toch zonder jullie moeten!

Lieve Victor, mijn grote broer. Wat een feestnummer ben jij! Altijd de grootste verhalen die stieken wat mooier zijn dan de werkelijkheid. Ik word er soms verlegen van hoe jij als grote broer opschept over wat je kleine zusje allemaal heeft gedaan of kan, met name als ze niet helemaal waarheidsgetrouw zijn ;).

Mijn lieve ouders, Marion en Nick. Door de stabiele thuissituatie die jullie jarenlang hebben gecreëerd en jullie onvoorwaardelijk steun en liefde heb ik me altijd een zondagskind gevoeld. Niets stond me in de weg en echte problemen bestonden eigenlijk niet. Helaas is aan onze ‘winning streak’ een einde gekomen met het grote verlies van Marion, maar ik weet zeker dat wij als familie ons daar kranig doorheen zullen slaan. Hopelijk is deze dag weer een voorzichtige start voor meer positievere dingen binnen ons gezin.

Lieve Rutger, woorden schieten te kort om uit te drukken hoe blij ik met je ben. Je bent mijn onvoorwaardelijke steun en toeverlaat. Alles is leuker en beter met z’n tweeën. Met jou aan mij zijde weet ik dat ik alles aan kan. Ik hou van je!
About the author
About the authors

Laura Lotte van Loendersloot was born in Laren, the Netherlands in 1982. She spent her childhood in Blaricum playing the saxophone, tennis, field hockey and singing in the kitchen with her mother.

In 2000 she graduated from Laar & Berg, and started medical school at the University of Amsterdam. Laura was a bright student; she effortlessly combined an active student life at LANX with her studies. In 2005, she wrote her first paper at the department of vascular medicine, under supervision of Prof. dr. Kastelein.

Being gifted with good hands skills and intelligence she decided to specialize in gynaecology. After her medical internships she joined the Center for Reproductive Medicine lead by Prof. dr. Van Der Veen for a PhD titled: Predicting IVF Outcome. During her PhD she also worked as a fertility doctor. To date, she has published 11 papers, gave two oral presentations and four posters presentations at international conferences and impregnated more women than the average man.

Currently, she is a gynaecological resident at the Sint Lucas Andreas Hospital in Amsterdam. Here she is board member of the residents association. She lives together with Rutger and their cat, Balou and will be expecting their first child in February. We have no doubt that our witty, smart, modest, bombshell friend will have a prosperous future.

Anna and Suthesh
Predicting IVF Outcome

On 25 July 1978 at 11:47 PM Louise Brown was born as the first IVF baby ever. Since its introduction more than 5 million babies have been born worldwide using IVF. In contrast to patients’ perception, IVF does not guarantee success; almost 50% of couples that start with IVF will not achieve a pregnancy through IVF even if they undergo multiple cycles. Given this limited success, it seems logical to offer IVF only to couples with reasonable chances of success and to discontinue treatment when chances are low and do not outweigh the burden and costs associated with treatment. As doctors are not able to correctly predict these pregnancy chances, prediction models can be a useful tool.

Another concern in current IVF practice is the high multiple pregnancy rates as multiple pregnancies are associated with an increase in maternal and perinatal morbidity and mortality as well as costs. A more individualized embryo transfer strategy could be a solution.

This PhD thesis describes the development and validation of several prediction models in IVF. The first part of this thesis focuses on couples’ prognosis with IVF. The second part of this thesis focuses on optimizing embryo transfer strategies.
Stellingen behorende bij het proefschrift

Predicting IVF Outcome

1. Het Templeton model kan maar matig onderscheid maken tussen paren met een hoge of lage zwangerschapskans met IVF, bovendien onderschat het model de zwangerschapskansen systematisch. Voor de huidige IVF praktijk heeft dit model geen toegevoegde waarde meer. *(dit proefschrift)*

2. De kans op een succesvolle IVF/ICSI behandeling wordt grotendeels bepaald door de leeftijd van de vrouw. *(dit proefschrift)*

3. Een multivariabel predictie model kan als hulpmiddel dienen in de spreekkamer om paren te adviseren bij de beslissing om al dan niet te starten of door te gaan met IVF. *(dit proefschrift)*

4. Het predictie model voor het rangschikken van embryo’s op basis van hun implantatie kans is in staat embryo’s met een hoge implantatie kans te onderscheiden van embryo’s met een matige of lage implantatie kans. *(dit proefschrift)*

5. IVF aanbieden aan paren met onverklaarde- of milde mannelijke subfertiliteit zonder gedegen wetenschappelijke onderbouwing is niet alleen zorgelijk vanuit een economisch maar ook vanuit een ethisch oogpunt.

6. Een tweeling zwangerschap na IVF/ICSI bij vrouwen boven de 38 is niet perse een complicatie.


8. “Science is the great antidote to the poison of enthusiasm and superstition”. *(Adam Smith)*

9. “Figures don’t lie, but liars figure”. *(Mark Twain)*

10. “There’s no evidence whatsoever that men are more rational than women. Both sexes seem to be equally irrational”. *(Albert Ellis)*

Laura van Loendersloot, december 2013