

About the (dis)continuation of fertility treatment

Citation for published version (APA):

Verhagen, T. E. M. (2014). *About the (dis)continuation of fertility treatment*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20140327tv>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/dis.20140327tv](https://doi.org/10.26481/dis.20140327tv)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

About the
(dis)continuation of
fertility treatment

About the (dis)continuation of fertility treatment

Tamara Elisabeth Maria Verhagen

Printing of this thesis was supported by:
Ferring B.V. and Hologic Benelux B.V.

About the (dis)continuation of fertility treatment
Thesis, University of Maastricht, The Netherlands

ISBN	978-94-6259-057-1
Cover and Lay-out	Casper Flipsen, VormMeester
Printed by	Ipskamp Drukkers

© copyright 2014, Tamara Verhagen

About the (dis)continuation of fertility treatment

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht
op gezag van de Rector Magnificus Prof. Dr. L.L.G. Soete
volgens het besluit van het College van Decanen
in het openbaar te verdedigen op
donderdag 27 maart 2014 om 10:00 uur.

door

Tamara Elisabeth Maria Verhagen

Promotiecommissie

Promotores

Prof. dr. J.L.H. Evers

Prof. dr. B.W. Mol (*University of Adelaide, Australia*)

Beoordelingscommissie

Prof. dr. C. de Die-Smulders, voorzitter

Prof. dr. D.D.M. Braat (*Radboudumc Nijmegen*)

Prof. dr. C.D. Dirksen

Prof. dr. F.M. Helmerhorst (*Leids Universitair Medisch Centrum*)

Prof. dr. H.C. Schouten

Contents

Chapter 1	General introduction and outline of the thesis.	7
Part I	Patient preferences	17
Chapter 2	The patients' perspective on the diagnostic fertility work-up. <i>Submitted</i>	19
Part II	Dropout	35
Chapter 3	Reasons for dropping out from a waiting list for in vitro fertilization. <i>Fertility Sterility 2010;94:1713-1716</i>	37
Chapter 4	What is the most accurate estimate of pregnancy rates in IVF dropouts? <i>Human Reproduction 2008;23:1793-1799</i>	49
Chapter 5	Impact of dropout during fertility treatment on the estimation of pregnancy rates in assisted reproduction. <i>Submitted</i>	65
Part III	Ovarian Reserve	81
Chapter 6	Value of ovarian reserve testing before IVF: a clinical decision analysis. <i>Human Reproduction 2006;21:1816-1823</i>	83
Chapter 7	The accuracy of multivariate models predicting ovarian reserve and pregnancy after in vitro fertilization: a meta-analysis. <i>Human Reproduction Update 2008;14:95-100</i>	101
Chapter 8	Summary and discussion	119
	Samenvatting	127
Appendices	Dankwoord	135
	Curriculum Vitae	139

Chapter 1

General introduction and
outline of the thesis

Introduction

In the general population 84% of couples will conceive within 1 year if the woman is aged less than 40 years and the couple has regular unprotected intercourse. Of those who do not conceive in the first year, about half will do so in the second year and cumulatively 92% of couples is pregnant after 2 years (te Velde *et al.*, 2000). Subfertility is usually defined as failure to conceive after 1 year of regular unprotected sexual intercourse (Zegers-Hochschild *et al.*, 2009) and it affects 9-15% of couples of childbearing age (Boivin *et al.*, 2007). Couples who have not conceived after 1 year, and are concerned about their fertility are eligible for clinical assessment and investigations: the diagnostic fertility work-up. About 56% of these couples will seek medical assistance to conceive (Boivin *et al.*, 2007).

After the diagnostic work-up couples can be diagnosed with unexplained subfertility (30%), ovulation disorders (20%), male subfertility (25%), tubal obstruction or defect or endometriosis (25%) (Collins and Van Steirteghem, 2004). These results are indicative for possible treatment options. When a couple's subfertility is unexplained, a validated prediction model is used to predict the chance of a spontaneous pregnancy through natural conception in the next 12 months (Hunault *et al.*, 2004; van der Steeg *et al.*, 2007). Expectant management is the best 'treatment' option when this chance is greater than or equal to 30% (Steures *et al.*, 2006), and optimal guideline adherence in intrauterine insemination (IUI) care has shown economical benefits according to cost-effective analysis in IUI care (Custers *et al.*, 2012; Haagen *et al.*, 2013).

When a couple has been unsuccessful in trying to get pregnant for at least one year, and visits a physician they will be confronted with a lot of information and will have a lot of decisions to make. Several studies have shown that these couples prefer informed decision making (Blenner, 1990, Stewart *et al.*, 2001; Dancet *et al.*, 2010; Dancet *et al.*, 2011). The first thing they have to decide is whether or not to start the diagnostic fertility work-up and at what moment to start it. After the fertility work-up, the results of the work-up will be discussed and an advice will be formulated. The couple makes a (informed) decision what to do with this information. Almost half of the couples decide not to start with fertility treatment. Brandes *et al.* (2009) reported that 6% of couples did not return after their first visit to the clinic, 3,4 % decided to quit during and 35,7% after the diagnostic fertility work-up. And many couples discontinue fertility treatment before a pregnancy is achieved, even when the prognosis of treatment is good. The world

live birth rate with assisted reproductive technologies (ART) is 22% per single initiated cycle of treatment (de Mouzon *et al.*, 2009). This can increase to 49% or higher if couples receive three or more cycles (Witsenburg *et al.*, 2005; Verhagen *et al.*, 2008). Reported dropout rates for IUI range from 4,5% to 60,9% (Dickey *et al.*, 2005; Brandes *et al.*, 2009; de Brucker *et al.*, 2009; Fahri *et al.*, 2010; Groeneveld *et al.*, 2011; Custers *et al.*, 2013) and for in vitro fertilization (IVF) from 7% to 69% (Sharma *et al.*, 2002; Pelinck *et al.*, 2006; Rajkhowa *et al.*, 2006; Brandes *et al.*, 2009; Malazia *et al.*, 2009).

There are several reasons for couples to discontinue treatment. The most important reason is achieving a pregnancy due to treatment, or spontaneously in between treatment cycles. Other reasons to quit treatment are psychological and physical burden, relational and personal problems, treatment rejection and organizational and clinic problems (Gameiro *et al.*, 2012). Financial issues can also be a reason to quit treatment, but even in countries where the costs of ART are fully covered by the national health system or by reimbursement from health insurance companies, couples decide to discontinue treatment (Domar *et al.*, 2004; Brandes *et al.*, 2009).

It is important to identify those subfertile couples having a good chance of a spontaneous pregnancy and couples who will benefit from fertility treatment. When subfertile couples with unexplained or mild male subfertility have an unfavourable prognosis of a spontaneous pregnancy, the first line treatment usually is IUI with mild ovarian stimulation for 6-9 cycles, and the second line usually is IVF (NVOG national guideline subfertility, 2011; NICE Guideline Fertility CG 156, 2013). IVF is a more expensive treatment with serious side effects, like ovarian hyperstimulation syndrome and multiple pregnancies. In IVF the occurrence of a poor response, often resulting in cancellation of the treatment cycle, is a common problem. Identification of potential poor responders before IVF treatment is commenced would potentially enable physicians to counsel these patients on the prospects of the treatment that they are about to begin, and even to actively withhold treatment in certain cases. A woman's age is the initial predictor of her chance of achieving a pregnancy through natural conception and with in vitro fertilisation (NICE CG 156, 2013). The following ovarian reserve tests (ORT) have been described to predict the likely ovarian response to gonadotrophin stimulation in IVF: Antral Follicle Count (AFC), Anti-Müllerian Hormone (AMH) serum level, and Follicle-Stimulating Hormone (FSH) serum level. A low total AFC, a low AMH level or a high basal FSH level are associated with a low ovarian response. The results of these ovarian reserve tests can be useful for the identification of low (and high) responders, in order to customize treatment. But ovarian reserve tests do not predict the chance of live birth or live full-term singleton

birth, which is the main treatment outcome of interest. Ovarian volume, ovarian blood flow, inhibin B serum level and oestradiol (E₂) serum level are other ovarian reserve tests that are evaluated. None of these tests do predict ovarian response following ovarian stimulation or outcome of fertility treatment. (NICE CG 156, 2013).

Background and outline of the present thesis

The fertility treatment of a particular couple is like a journey without itinerary: at one moment the couple starts with this journey, but the roads they will travel, the time that will pass and their final destination, achieving a pregnancy or the ultimate goal a healthy child, is uncertain. A lot of couples suspend this journey before the desired endpoint has been reached.

In this thesis we evaluate several parts of the fertility treatment subfertile couples will experience, when seeking medical advice and treatment for their fertility problem.

Couples not being pregnant after 1 year of unprotected intercourse can be referred for the diagnostic fertility work-up. In **chapter 2** we report the results from interviews with couples referred for the diagnostic fertility work-up: what is the reason for couples to start this work-up and which health risks are they prepared to accept during the work-up?

Not all couples on a waiting list for IVF actually start treatment. In **chapter 3** the incidence of dropout from an IVF waiting list is determined and the reasons for dropping out are described.

After starting IVF-treatment discontinuation of treatment is also a common problem, and dropout influences the reported (cumulative) pregnancy rates in ART studies. In **chapter 4** follow-up data of IVF dropouts are collected to calculate cumulative pregnancy rates for the couples continuing IVF-treatment and for the 'completed IVF-group', including the follow-up data from dropouts.

In **chapter 5** we present data on pregnancy rates and dropout rates in consecutive IUI and IVF treatment cycles from studies collected with a literature search. These data are analysed to find out whether there is a relation between a reported pregnancy rate and the dropout rate in the preceding cycle.

In **chapter 6** the results of a clinical decision analysis concerning the value of ovarian reserve testing are presented. We use a decision analytic framework, modelling the ovarian reserve test accuracy as well as patients and doctors perception of the outcomes. A group of patients and doctors working in the field of infertility are asked to judge about the amount of distress caused by both incorrectly starting IVF (i.e. IVF without a pregnancy) and incorrectly withholding IVF (i.e. not starting IVF where pregnancy would have occurred).

In **chapter 7** a meta-analysis is performed to assess the accuracy of multivariable models predicting ovarian reserve.

References

- Blenner JL. Attaining self-care in infertility treatment. *Appl Nurs Res* 1990; 3:98–104.
- Boivin J, Bunting L, Collins JA and Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;22:1506-1512.
- Brandes M, van der Steen JOM, Bokdam SB, Hamilton CJCM, de Bruin JP, Nelen WLDM, Kremer JAM. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod* 2009;24:3127–3134.
- Collins JA and Van Steirteghem A. Overall prognosis with current treatment of infertility. *Hum Reprod Update* 2004;10:309-316.
- Custers IM, van Rumste MME, van der Steeg JW, van Wely M, Hompes PGA, Bossuyt P, Broekmans FJ, Renckens CNM, Eijkemans MJC, van Dessel TJHM, van der Veen F, Mol BW, Steures P and CECERM. Long-term outcome in couples with unexplained subfertility and an intermediate prognosis initially randomized between expectant management and immediate treatment. *Hum Reprod* 2012;27:444-450.
- Custers IM, van Dessel TH, Flierman PA, Steures P, van Wely M, van der Veen F, Mol BW. Couples dropping out of a reimbursed intrauterine insemination program: what is their prognostic profile and why do they drop out? *Fertil Steril* 2013;99:1294-1298.
- Dancet EA, Nelen WL, Sermeus W, De Leeuw L, Kremer JA, D'Hooghe TM. The patients' perspective on fertility care: a systematic review. *Hum Reprod Update*. 2010 Sep-Oct;16(5):467-87.
- Dancet EA, Van Empel IW, Rober P, Nelen WL, Kremer JA, D'Hooghe TM. Patient-centred infertility care: a qualitative study to listen to the patient's voice. *Hum Reprod*. 2011 Apr;26(4):827-33.
- de Brucker M, Haentjes P, Evenepoel J, Devroey P, Collins J, Tounaye H. Cumulative delivery rates in different age groups after artificial insemination with donor sperm. *Hum Reprod* 2009;24:1891-1899.
- de Mouzon J, Lancaster P, Nygren KC, Sullivan E, Zegers-Hochschild F, Mansour R, Ishihara O, Adamson D. World collaborative report on assisted reproductive technology, 2002. *Hum Reprod* 2009;24:2310–2320.
- Dickey RP, Taylor SN, Lu PY, Sartor BM, Rye PH and Pyrzak R. Risk factors for high-order multiple pregnancy and multiple birth after controlled ovarian hyperstimulation: results of 4,062 intrauterine insemination cycles. *Fertil Steril* 2005;83:671-683.
- Domar AD. Impact of psychological factors on dropout rates in insured infertility patients. *Fertil Steril* 2004;81:271–273.
- Fahri J and Orvieto R. Cumulative clinical pregnancy rates after COH and IUI in subfertile couples. *Gynecol Endocrinol* 2010;26:500-504.

Gameiro S, Boivin J, Peronace L and Verhaak CM. Why do patients discontinue fertility treatment? A systematic review of reasons and predictors of discontinuation in fertility treatment. *Hum Reprod* 2012;18:652-669.

Groeneveld E, Kouijzer IJ, Timmermans AJ, Schats R and Hompes PG. Effectiveness of highly purified human menopausal gonadotrophin in intrauterine insemination. *Eur J Obstet Gynecol Reprod Biol* 2011;154:182-186.

Haagen EC, Nelen WLDM, Adang EM, Grol RPTM, Hermens RPMG, Kremer JAM. Guideline adherence is worth the effort: a cost-effectiveness analysis in intrauterine insemination care. *Hum Reprod* 2013;28:357-366.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH, te Velde ER. Two new prediction rules for spontaneous pregnancy leading to life birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;19:2019-2026.

Malizia BA, Hacker MR and Penzias AS. Cumulative live-birth rates after in vitro fertilization. *N Engl J Med* 2009;360:236-243.

NICE Clinical Guidelines 156. Guideline fertility: assessment and treatment for people with fertility problems. <http://guidance.nice.org.uk/CG156>

NVOG national guideline subfertility 2011. http://nvogdocumenten.nl/index.php?pagina=/richtlijn/pagina.php&fSelectTG_62=75&fSelectedSub=62&fSelectedParent=75

Pelinck MJ, Vogel NE, Hoek A, Simons AH, Arts EG, Mochtar MH, Beemsterboer S, Hondelink MN and Heineman MJ. Cumulative pregnancy rates after three cycles of minimal stimulation IVF and results according to subfertility diagnosis: a multicentre cohort study. *Hum Reprod*. 2006;21:2375-2383.

Rajkhowa M, McConnell A, Thomas GE. Reasons for discontinuity of IVF treatment: a questionnaire study. *Hum Reprod* 2006;21:358-363.

Sharma V, Allgar V, and Rajkhowa M. Factors influencing the cumulative conception rate and discontinuation of in vitro fertilization treatment for infertility. *Fertil Steril* 2002;78:40-46.

Steures P, van der Steeg JW, Hompes PG, Habbema JD, Eijkemans MJ, Broekmans FJ, Verhoeve HR, Bossuyt PM, van der Veen F, Mol BW for the Collaborative Effort on the Clinical Evaluation in Reproductive Medicine. Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 2006;368:216-221.

Stewart L, Hamilton M, McTavish A, Fitzmaurice A, Graham W. Randomized controlled trial comparing couple satisfaction with appointment and telephone follow-up consultation after unsuccessful IVF/ICSI treatment. *Hum Fertil (Camb)* 2001a;4:249-255.

te Velde ER, Eijkemans R, Habbema HDF. Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. *Lancet* 2000;355:1928-9.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ, Bossuyt PM, van der Veen F, Mol BW; CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine). Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod.* 2007 Feb;22(2):536-42.

Verhagen TEM, Dumoulin JCM, Evers JLH, Land JA. What is the most accurate estimate of pregnancy rates in IVF dropouts? *Hum Reprod* 2008;23:1793–1799.

Witsenburg C, Dieben S, Van der Westerlaken L, BVerburg H, Naaktgeboren N. Cumulative live birth rates in cohorts of patients treated with in vitro fertilization or intracytoplasmic sperm injection. *Fertil Steril* 2005;84:99–107.

Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E and Vanderpoel S. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril* 2009;92:1520-1524.

PART I

Patient
preferences

Chapter 2

The patients' perspective on the diagnostic fertility work-up

Tamara E.M. Verhagen

Femke Y. Mooren-Luttjeboer

Eline A. Dancet

Brent C. Opmeer

Sjors F. Coppus

Ben W. Mol

Submitted

Abstract

Little is known of patients' perspectives on diagnostic testing in the initial fertility work-up. We administered structured face-to-face interviews to subfertile couples prior to the diagnostic fertility work-up to assess their knowledge, concerns and desire for information and expectations regarding their role in the work-up. The interview included closed questions rated on Likert and visual analogue scales and trade-off questions. Data were analyzed with descriptive statistics. In total, 38 women and 28 of their male partners participated. Most patients assessed their knowledge as limited, and wanted to know as much as possible about the diagnostic work-up and fertility treatment. Almost all patients wanted to be involved in shared decision-making in the fertility work-up. Finding a cause for their subfertility was valued more than conceiving without knowing the cause ($p < 0.001$). Patients accept a risk of 1% on a complication during the work-up, described as perforation or infection at tubal testing. We conclude that subfertile patients want shared decision-making. To establish that, they require information and their perspective on the acceptable risk of a complication and acceptable rates of detecting a cause should be taken into account.

Introduction

Couples without a history of predisposing factors influencing fertility, are considered to be subfertile after at least 12 months of regular unprotected intercourse without conceiving (Zegers-Hochschild *et al.*, 2009). These couples are eligible for clinical assessment and investigations: the diagnostic fertility work-up and are referred to a gynaecologist. During the diagnostic fertility work-up, a medical history is taken and vital reproductive functions are assessed by means of blood tests, cycle evaluation with ovulation detection, semen analysis and assessment of the Fallopian tubes (NICE Guideline 11; Dutch society of Obstetrics and Gynaecology, 2010). In 75-80% of couples a cause for fertility problems can be established (Hull *et al.*, 1985; Cahill *et al.*, 2002). The risk on a complication during the investigations is 1-4%, mainly during testing of tubal patency where pain, uterine perforation or infection can occur (Rice *et al.*, 1986; Forsey *et al.*, 1990). Once assessment and investigations have been undertaken, a management plan can then be established with the couple. When no fertility problems are established, the term 'unexplained subfertility' is used, and, according to the Dutch guidelines, the probability of a spontaneous ongoing pregnancy within the next year is calculated using a validated prediction model (Dutch Society of Obstetrics and Gynaecology, 2010; Steures *et al.*, 2006; van der Steeg *et al.*, 2007). The calculated probability directs the therapy proposed to the unexplained subfertile couple; for example if the prediction model predicts a chance of a spontaneous pregnancy in the next year equal to or greater than 30%, expectant management for six months is advised.

Not all subfertile couples will seek medical assistance to conceive (Boivin *et al.*, 2007). Once referred to a gynaecologist, couples should be counselled about the assessments and investigations done during the diagnostic fertility work-up and the possible outcomes. Before starting the work-up they can decide to do nothing yet; i.e. expectant management, and postpone the work-up. A study of Brandes *et al.* showed that 45% of subfertile couples that contacted a fertility clinic discontinued fertility care before they started any treatment, of which almost 10% even after the first visit or during the work-up (Brandes *et al.*, 2009). Subfertile couples value information provision and the possibility of shared decision-making (Blenner, 1990, Stewart *et al.*, 2001; Dancet *et al.*, 2010; Dancet *et al.*, 2011; Dancet *et al.*, 2012). The rare studies on perspectives of subfertile couples on fertility investigations (Dancet *et al.*, 2010) had contradictory results on the sufficiency of information provision (Culley *et al.*, 2006; Souter *et al.*, 1998) and reported negative experiences regarding excessive repetition of tests and excessive waiting times for investigations communication of test results (Souter *et al.*, 1998). A study on women's

perspectives on hysterosalpingography (HSG) and hysterosalpingo-contrastsonography (HyCoSy) did not find differences in pain and anxiety and found that, after experiencing both procedures, most women preferred one of the tests, however this could be HSG (41%) or HyCoSy (34%) (Dijkman *et al.*, 2000).

To our knowledge, the decision between expectant management or the diagnostic fertility work-up that subfertile couples make, has not been explored yet. Physicians should be aware of the couples' need for information and their trade-offs between advantages and disadvantages of investigations in the work-up (Kassirer, 1994; Haynes *et al.*, 2002). Guidelines presume that patients accept a small risk of a complication but empirical evidence for this presumption is missing.

We performed this study to document concerns, knowledge, desire for shared decision-making and trade-offs of safety and effectiveness regarding the decision between expectant management and the diagnostic fertility work-up.

Material and methods

Patient recruitment

Subfertile couples referred to the Máxima Medical Centre for a fertility work-up were invited to participate in the study. Couples were eligible if they had sufficient command of the Dutch language and if they had been subfertile for at least one year. Prior to their first consultation, eligible couples were sent written study information by mail and were subsequently contacted by phone by a researcher (FL or TV) who explained the study objective and procedure. Although couples were invited, women were interviewed separately if their partner was not willing to participate. The interview was scheduled before the results of the fertility work-up were discussed with the couples.

Questionnaire administration

A face-to-face paper-pencil questionnaire administration was scheduled prior to or at the beginning of the fertility work-up. After obtaining patient's written informed consent, the questionnaire was administered by a researcher (TV or FML) at the hospital during a 45-minute session. Both researchers followed the same study protocol. Questions were read out loud by the researchers and patients could read the questions again themselves before answering them on their copy of the questionnaire. Couples were interviewed in the same session but each partner filled his/ her own questionnaire. If required, the researcher gave additional explanations.

The questionnaire

The questionnaire consisted of five parts:

Part one included a limited set of demographic and medical background questions.

Part two addressed couples' desire for information about the advantages and disadvantages of the diagnostic fertility work-up on a visual analogue scale (VAS) with '0' representing 'as much information as possible' and '10' representing 'as little information as possible'. Participants were also asked whether they preferred to be informed extensively about the diagnostic options to make their own decisions (VAS=0), or not to be informed and let the doctor decide what to do (VAS=10).

Part three encompassed the following topics with the aid of 5-point Likert scales: (I) concerns and desire of knowledge of the diagnostic fertility work-up, fertility treatment, decreased fertility and possible childlessness, (II) self-assessment of personal knowledge on the diagnostic fertility work-up and therapeutic options.

The answering categories of the Likert scales were 'not at all' (1), 'a little' (2), 'neutral' (3), 'quite a lot' (4) or 'extreme' (5).

Part four questioned the couples' motivations for the diagnostic fertility work-up and their expectations regarding the factor causing their subfertility. Motivations for the work-up were "I want to know why we do not get pregnant" or "I want to become pregnant" or both. A couple of assertions concerning these motivations were formulated and the participants were asked to weight each assertion on a visual analogue scale (0 = not important; 10 = very important). Thereafter we asked the participants to estimate the possibility of finding a causal factor for their infertility during the fertility work-up. They had to indicate whether they thought a causal factor was found in themselves, in their partner, in both or in none of them (unexplained) respectively by scoring a percentage ranging from 0-100%.

Part five of the interview was designed to explore the trade-offs that would be made by patients if they were involved in decision-making on the diagnostic dilemma. Therefore, the interview included a graph on the diagnostic dilemma (figure 1). Patients were asked to choose between 'immediate diagnostic fertility work-up' or '6 months of expectant management' (i.e. postponing the diagnostic fertility work-up while the spontaneous pregnancy rate would be 40%), while varying the 'risk rate' of the diagnostic fertility

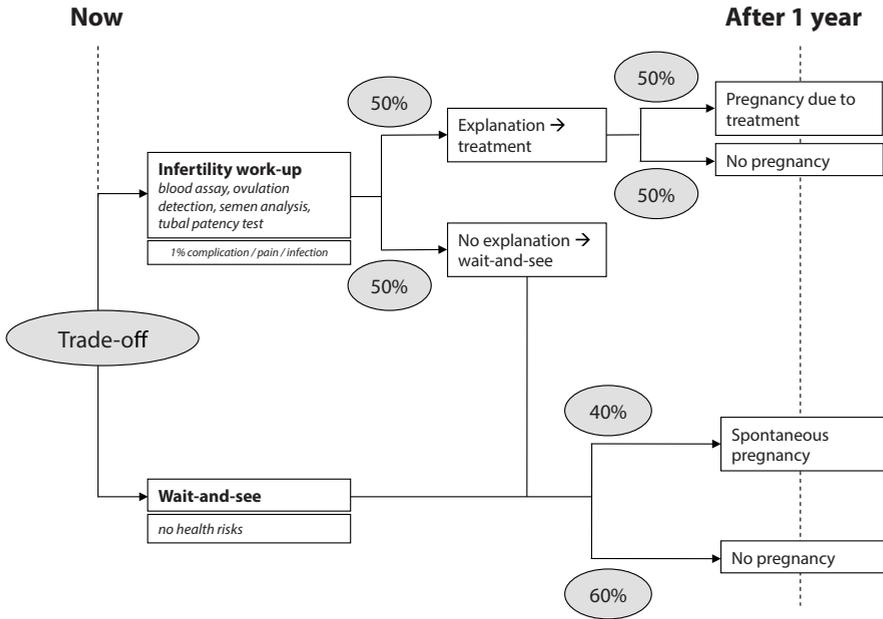


FIGURE 1. Trade-off: infertility work-up versus wait-and-see policy. The complication risk (initially set at 1%) and the chance of detecting an abnormality during the fertility work-up (initial set at 50%) were varied in the trade-off.

work-up (i.e. risk on severe pain, an infection or uterine perforation as complication of the work up) and the ‘effectiveness rate’ (i.e. likeliness of detecting an explanation for the fertility problem). Regarding the varying ‘risk rates’ and ‘effectiveness rates’ of the diagnostic fertility work-up, the ‘effectiveness rate’ was initially set at 50% and the ‘risk rate’ was initially set at 1%. In case patients initially choose ‘immediate diagnostic fertility work-up’, the ‘risk rate’ was first increased by steps of 1, 5 and 10% until their preference switched to ‘six months of expectant management’, while the ‘effectiveness rate’ was 50%. Second, the ‘effectiveness rate’ was decreased in steps of 1, 5 and 10% until their preference switched to ‘six months of expectant management’, while the ‘risk rate’ was 1%.

Similarly, when a patient initially preferred ‘six months of expectant management’, the ‘risk rate’ was decreased first and the ‘effectiveness rate’ was raised second until the preference switched.

Data analysis

Data were entered in the Statistical Package for Social Sciences (SPSS) version 13.0 and analyzed with descriptive statistics and exploratory data analysis focusing on differences between male and female participants and between questions.

For presentation and analysis the answers on the 5-point Likert scales were dichotomized (i.e. the scales were collapsed into a positive or negative answers, with the neutral response was added to the negative assessment group; Dancet *et al.*, 2010). Comparing the dichotomized answers on the Likert-scales of different groups or on different questions was done with Chi-square tests or Fishers exact test. For VAS-scales median and ranges were computed and compared with Mann-Whitney U test.

With respect to the answers to the trade-off questions the following outcomes were computed: the initial preferences distribution, the mean 'risk rate' at which patients changed their preference, the minimal and maximal 'risk rate' accepted by a patient, the 'risk rate' accepted by the majority of patients; the mean 'effectiveness rate' at which patients changed their preference; the minimal and maximal 'effectiveness rate' accepted by a patient and the 'effectiveness rate' accepted by the majority of patients.

Results

The participants

There were 126 referrals for fertility care, 20 couples were excluded because of an obvious known reason for not conceiving (anovulation or male infertility), being referred before they were officially subfertile (referred within 12 months) or not able to read and speak the Dutch language. From the 106 eligible couples, 38 (36%) was willing to participate. We were able to interview all women and 28 of their male partners. Partners who were not willing to participate mainly did so because of time constraints. Reasons for couples not to participate were: a spontaneous pregnancy before the first visit in six couples, time constraints in 22 couples, and no particular mentioned reason in the other couples. It took 17 months to select these couples and conduct the interviews. The mean age for female participants was 31.5 years (range 22-37) and 33.0 years (range 23-42) for male participants. In 33 couples subfertility was primary and in five couples it was secondary. Three of the five couples with secondary subfertility already had a child. The mean duration of subfertility was 18 months (range 12-36 months).

Patients' motivation for the fertility work-up and their desire for informed decision-making

Finding an explanation for not conceiving was a more important motivation for participants to enrol in the diagnostic fertility work-up than achieving a conception (respective median VAS scores of 8 and 5; $p < 0.001$). Men and women were comparable in this respect.

Sixty-three participants (95%) preferred shared decision-making on the diagnostic fertility work-up. The paternalistic decision-making approach, in which the physician decided without consulting the patient, was preferred by none of the patients. Women and men did not differ in their desire for informed and shared decision-making.

Patients' desire for and self-assessment of knowledge for the fertility work-up

Figure 2a reports on the participants' concerns and desire of knowledge of the diagnostic fertility work-up, fertility treatment, decreased fertility and possible childlessness. No differences were observed for men and women. Figure 2b shows the participants' self-assessment of personal knowledge of the diagnostic fertility work-up, fertility treatment and the possible consequences of subfertility and fertility treatment. Women reported a higher level of knowledge about the fertility work-up than men (16 women (42%) versus 5 men (18%); $p = 0.04$).

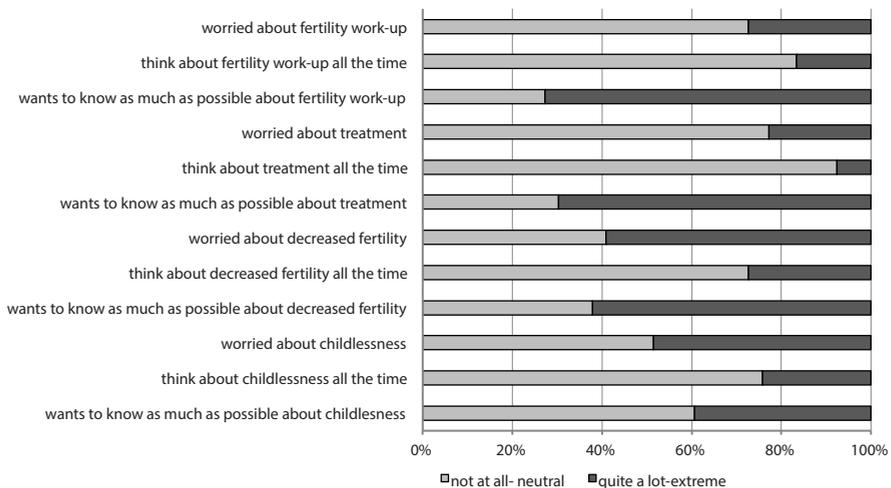


FIGURE 2A. Participants' valuation of their concerns and desire of knowledge of the fertility work-up, fertility treatment, decreased fertility and possible childlessness.

Women compared to men knew more often couples with fertility problems (19 women (50%) versus 7 men (25%); $p=0.04$), read more information about the fertility work-up (12 women (32%) versus 1 men (4%); $p=0.005$) and read more about the consequences of subfertility (7 women (18%) versus 0 men; $p=0.02$).

Patients' assumptions on the cause of their fertility problem

Women deemed it more likely than men that a possible explanation for their subfertility was found in themselves (65% versus 40%; $p=0.03$). Men deemed it more likely than women that the possible explanation for their subfertility was found in their partner (55% versus 30%; $p=0.04$). Women and men deemed it equally likely that no cause of subfertility would be found (respectively 40% and 36%; $p=0.56$).

Patients' preference in the trade-off

Fifty-three participants (82%) initially preferred the diagnostic fertility work-up, whereas nine (14%) preferred six months of expectant management and three (4%) did not express a preference. The preference of the participants who initially had chosen for the diagnostic fertility work-up, switched from diagnostic fertility work-up to six months of expectant management at a mean 'risk rate' of 20%; respectively 23% for women and 15% for men ($p=0.22$). The minimal accepted 'risk rate' was 2% and two women would even choose for the fertility work-up when the chance on a complication was 100%.

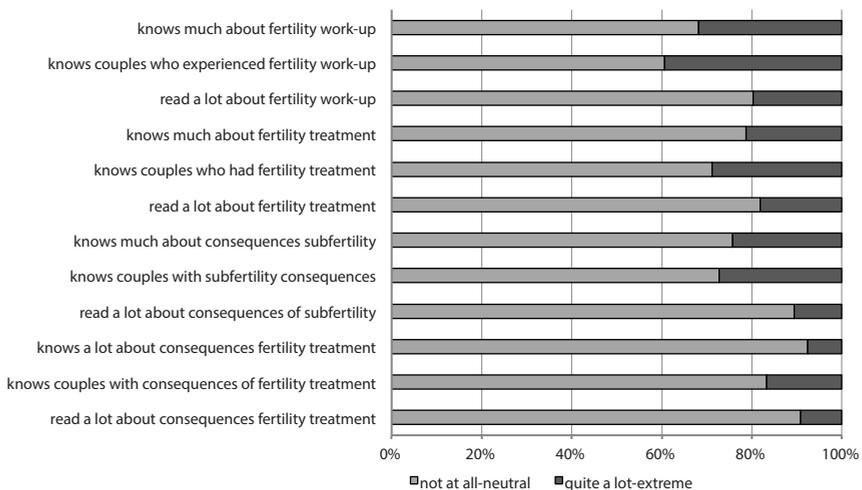


FIGURE 2B. Participants' self-assessment of their personal knowledge of the diagnostic fertility work-up, fertility treatment, the consequences of subfertility and the consequences of fertility treatment.

When the 'risk rate' of the diagnostic fertility work-up exceeded 10%, 50% of participants preferred a six months of expectant management (Figure 3).

The mean 'effectiveness rate' at which participants switched preference from diagnostic fertility work-up to six months of expectant management was 24%; respectively 26% for women and 21% for men ($p=0.35$). Three participants were persistent in their choice for the diagnostic fertility work-up even when the 'effectiveness rate' of the diagnostic fertility work-up was 0%. The majority of participants (50%) would prefer six months of expectant management if the chance of detecting a cause during the fertility work-up was less than 25% (Figure 4).

Discussion

To our knowledge this is the first study on subfertile couples' perspectives on the decision that has to be made between expectant management and diagnostic fertility work-up, as most studies have focused on perspectives of couples starting or already undergoing treatment (Schmidt *et al.*, 2003; Dancet *et al.*, 2010). In this study, we investigated the perspective on initial fertility care of subfertile couples before they start with fertility treatment. This is an important group as almost half of the couples discontinuing fertility care do so before treatment even starts (Brandes *et al.*, 2009). This might indicate that couples at first are more interested in finding a cause of their subfertility than conceiving as soon as possible, when they are embarking on fertility care. And this is consistent with the results in our study: participants wanted to know whether there is an explanation for not conceiving yet rather than conceiving as soon as possible. It might be possible that the preference will be different at the end of the diagnostic work-up, after having discussed the results with a physician. A Danish study among infertile couples has shown that almost all their participants stated that achieving a pregnancy was the most important reason to seek treatment and only 56% was interested in finding a cause for not conceiving. These couples were infertile for on average 4.1 years and most of these couples were about to start IVF treatment (Schmidt *et al.*, 2003). Other preference studies in fertility treatment showed that preference might switch when time passes by. Van Weert *et al* found that couples undergoing intrauterine insemination (IUI) preferred continuing IUI over in vitro fertilization (IVF) after three cycles and after six cycles the preference switched to IVF treatment (van Weert *et al.*, 2007).

The present study has several other interesting findings that add to previous knowledge.

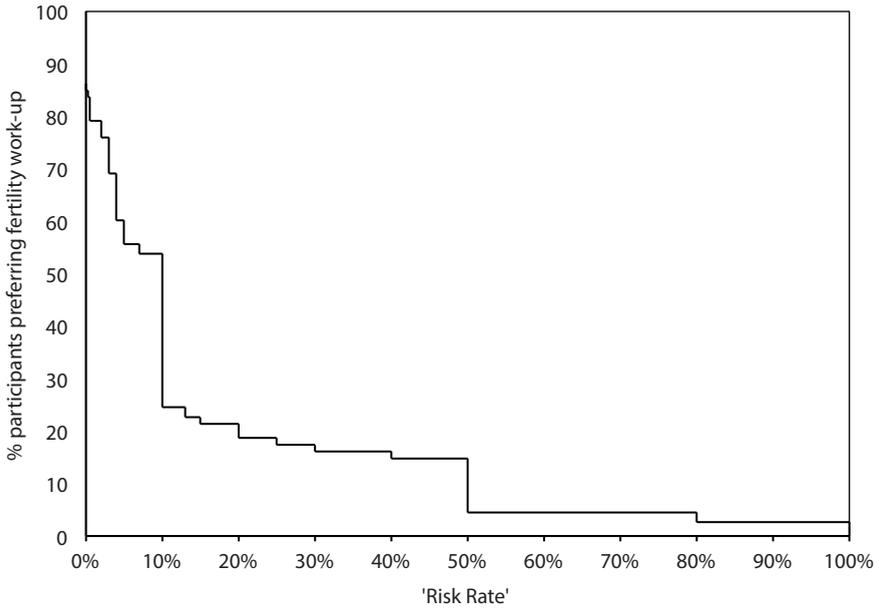


FIGURE 3. Participants' thresholds for the probability of a complication during the diagnostic fertility work-up (= the 'Risk Rate')

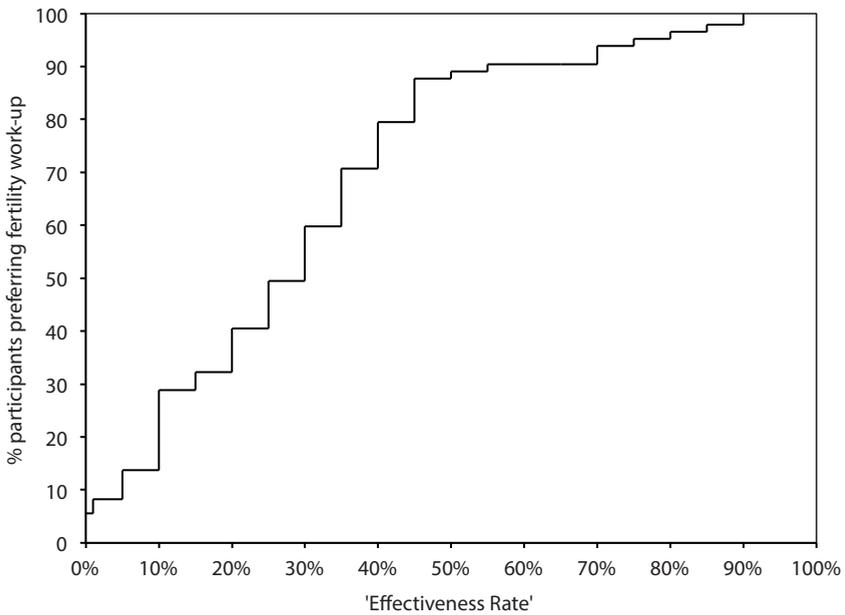


FIGURE 4. Participants' thresholds for variable probabilities of finding an explanation for their subfertility during the diagnostic fertility work-up (= the 'Effectiveness Rate').

In a hypothetical scenario, most of the participants in this study reported to accept a 1% complication risk assuming there would be a 50% chance on finding an explanation for their subfertility. In the trade-off both women and men were willing to take a 20% complication risk on average assuming there would be a 50% chance on finding an explanation for their subfertility. Only a small proportion of patients preferred the wait-and-see policy and would postpone the diagnostic fertility work-up. Couples interviewed in our study already had decided to start the fertility work-up: they consulted their general practitioner because they did not conceive and were referred to the hospital for further examinations. This may have caused selection bias because couples not willing to embark fertility treatment or even examinations in a diagnostic fertility work-up will not visit a general practitioner. Nevertheless, it was obvious during the interviews that not all participants were aware of the examinations that were part of the fertility work-up and for that reason the interviews gave us the information we wanted. Moreover, to interview only couples even before they consulted a physician was not feasible in our study.

Implications for daily practice

Our study showed that couples preferred to be informed as much as possible and almost all of them preferred the informed decision-making strategy in which they can choose after being informed by their physician. Daily practice often implicates that physicians tell patients what to do when the results of the diagnostic fertility work-up are discussed. Previous preference research by Steures *et al.* showed that the majority of subfertile couples preferred intrauterine insemination with or without controlled ovarian hyperstimulation when the chance of a spontaneous pregnancy in the next 12 months was <40% and <50% respectively (Steures *et al.*, 2005). In the Netherlands couples are told to wait for a spontaneous pregnancy for 6-12 months when their chance of a spontaneous pregnancy is 30% or more in the next 12 months calculated by a validated prediction model (Hunault *et al.*, 2004; van der Steeg *et al.*, 2007). Besides not achieving more ongoing pregnancies with treatment (i.e. IUI with controlled ovarian stimulation) over expectant management in these couples (Steures *et al.*, 2006), optimal guideline adherence in IUI care has economical benefits according to cost-effective analysis in IUI care (Custers *et al.*, 2012; Haagen *et al.*, 2013). These important aspects of the benefits of expectant management must be part of counselling couples about the importance of expectant management in fertility care.

Another important issue in counselling subfertile couples is to be aware of the possible differences between women and men. In our study women reported a higher level of

knowledge of the fertility work-up than men (figure 2b). This can be explained by the fact that women compared to men more often knew couples with fertility problems, read more information about the fertility work-up and read more about the consequences of subfertility. This might confirm that men's involvement and sense of partnership in fertility care is only partially accomplished, as also has been shown in other studies where men called themselves 'onlookers' rather than feeling involved in fertility care (Hinton and Miller, 2013).

Evidence-based medicine aims to integrate the best available evidence with clinical expertise, and patient values (Sackett *et al.*, 1996). At present, Dutch guidelines for the diagnostic fertility work-up of subfertile couples recommend to start with a diagnostic fertility work-up (i.e. cycle monitoring, basal blood test and semen analysis) and advise an expectant management in case the chance of a spontaneous pregnancy within the next 6 months is realistic. When no abnormality is expected tubal testing will be postponed. Both women and men state that a risk of 20 % is acceptable for a 50% chance to find an explanation. Tubal patency tests could be done earlier in the work-up according to the interviewed subfertile couples in the Máxima Medical Centre.

Conclusions

Subfertile couples consulting a specialist for subfertility want to be well informed on the nature and purpose of diagnostic procedures during the diagnostic fertility work-up. They claim full participation in shared decision-making on the diagnostic steps. They can indicate their individual threshold of acceptance of complication risks and prognosis in the process of making choices. Gaining insight on the cause of their subfertility seems more important than conceiving without knowing the cause.

References

- Blenner JL. Attaining self-care in infertility treatment. *Appl Nurs Res* 1990; 3:98–104.
- Boivin J, Bunting L, Collins JA and Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;22:1506-1512.
- Brandes M, van der Steen JOM, Bokdam SB, Hamilton CJCM, de Bruin JP, Nelen WLDM, Kremer JAM. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod* 2009; 24:3127-3135.
- Cahill DJ, Wardle PG. Management of infertility. *BMJ* 2002;325 (7354):28-32.
- Culley LA, Hudson N, Rapport FL, Katbamna S, Johnson MRD. British South Asian communities and infertility services. *Hum Fertil (Camb)* 2006;9:37–45.
- Custers IM, van Rumste MME, van der Steeg JW, van Wely M, Hompes PGA, Bossuyt P, Broekmans FJ, Renckens CNM, Eijkemans MJC, van Dessel TJHM, van der Veen F, Mol BW, Steures P and CECERM. Long-term outcome in couples with unexplained subfertility and an intermediate prognosis initially randomized between expectant management and immediate treatment. *Hum Reprod* 2012;27:444-450.
- Dancet EA, Nelen WL, Sermeus W, De Leeuw L, Kremer JA, D’Hooghe TM. The patients’ perspective on fertility care: a systematic review. *Hum Reprod Update*. 2010 Sep-Oct;16(5):467-87.
- Dancet EA, Van Empel IW, Rober P, Nelen WL, Kremer JA, D’Hooghe TM. Patient-centred infertility care: a qualitative study to listen to the patient’s voice. *Hum Reprod*. 2011 Apr;26(4):827-33.
- Dancet EA, D’Hooghe TM, Sermeus W, van Empel I, Strohmer H, Wyns C, Santa-Cruz D, Nardo LG, Kovatchki D, Vanlangenakker L, Garcia-Velasco J, Mulugeta B, Nelen WL, Kremer JA. Patients from across Europe have similar views on patient-centered care: an international multilingual qualitative study in infertility care. *Hum Reprod*. 2012 Jun;27(6):1702-11.
- Dijkman AB, Mol BWJ, van der Veen F, Bossuyt PMM, Hogerzeil HV. Can hysterosalpingocontrast-sonography replace hysterosalpingography in the assessment of tubal subfertility. *Eur J Radiol* 2000;35:44–48.
- Dutch Society of Obstetrics and Gynaecology (NVOG). Nationwide Network Guideline Subfertility 2010.
- Forsey JP, Caul EO, Paul ID, Hull MG. Chlamydia trachomatis, tubal disease and the incidence of symptomatic and asymptomatic infection following hysterosalpingography. *Hum Reprod* 1990;5:444-7.
- Haagen EC, Nelen WLDM, Adang EM, Grol RPTM, Hermens RPMG, Kremer JAM. Guideline adherence is worth the effort: a cost-effectiveness analysis in intrauterine insemination care. *Hum Reprod* 2013;28:357-366.
- Haynes RB, Devereaux PJ, Guyatt GH. Physicians’ and patients’ choices in evidence based practice. *BMJ* 2002;234:1350.

Hinton L, Miller T. Mapping men's anticipations and experiences in the reproductive realm: (in) fertility journeys. *Reprod Biomed Online* 2013;27:244-252.

Kassirer JP. Incorporating patients' preferences into medical decisions. *N Engl J Med.* 1994 Jun 30;330(26):1895-6.

NICE Clinical Guideline no. 11. Fertility: assessment and treatment for people with fertility problems. February 2004. <http://www.nice.org.uk/Guidance/CG11/NiceGuidance/pdf/English>

Rice JP, London SN, Olive DL. Reevaluation of hysterosalpingography in infertility investigation. *Obstet Gynecol* 1986;65:718-721.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.

Schmidt L, Holstein BW, Boivin J, H.SangrenH, Tjørnhøj-Thomsen T, J.Blaabjerg J, Hald F, Nyboe Andersen A and Rasmussen PE. Patients' attitudes to medical and psychosocial aspects of care in fertility clinics: Findings from the Copenhagen Multi-centre Psychosocial Infertility (COMPI) Research Programme. *Human Reprod.* 2003;18(3):628-637.

Souter VL, Penney G, Hopton JL, Templeton AA. Patient satisfaction with the management of infertility. *Hum Reprod.* 1998;13(7):1831-6.

Steures P, Berkhout JC, Hompes PG, van der Steeg JW, Bossuyt PM, van der Veen F, Habbema JD, Eijkemans MJ, Mol BW. Patients' preferences in deciding between intrauterine insemination and expectant management. *Hum Reprod.* 2005;20(3):752-5.

Steures P, van der Steeg JW, Hompes PG, Habbema JD, Eijkemans MJ, Broekmans FJ, Verhoeve HR, Bossuyt PM, van der Veen F, Mol BW for the Collaborative Effort on the Clinical Evaluation in Reproductive Medicine. Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 2006;368:216-221.

Stewart L, Hamilton M, McTavish A, Fitzmaurice A, Graham W. Randomized controlled trial comparing couple satisfaction with appointment and telephone follow-up consultation after unsuccessful IVF/ICSI treatment. *Hum Fertil (Camb)* 2001a;4:249-255.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ, Bossuyt PM, van der Veen F, Mol BW; CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine). Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod.* 2007 Feb;22(2):536-42.

Van Weert JM, van den Broek J, van der Steeg JW, van der Veen F, Flierman PA, Mol BW, Steures P. Patients' preferences for intrauterine insemination or in-vitro fertilization. *Reprod Biomed Online* 2007;15(4):422-427.

Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E and Vanderpoel S. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril* 2009;92:1520-1524.

PART II

Dropout

Chapter 3

Reasons for dropping out from a waiting list for in vitro fertilization

Angelique J.C.M. van Dongen

Tamara E.M. Verhagen

John C.M. Dumoulin

Jolande A. Land

Johannes L.H. Evers

Fertility and Sterility 2010; 94:1713-1716

Abstract

OBJECTIVE: To determine the incidence of couples dropping out of the in vitro fertilization (IVF) waiting list and to describe the couples' reasons.

DESIGN: Prospective cohort study.

SETTING: Fertility centre in an academic hospital.

PATIENTS: 674 women placed consecutively on the IVF waiting list between June 2000 and July 2003.

INTERVENTIONS: None.

MAIN OUTCOME MEASURES: Number of dropouts and reasons for dropping out.

RESULTS: Follow-up information was collected in 2005 and 2008. Of the 674 couples on the waiting list, 87% started IVF, and 13% dropped out before starting their first IVF cycle. Follow-up data were obtained for 85/86 patients (98.8%): 37% dropped out because of spontaneous pregnancy, 36% for personal reasons (passive censoring), and 27% for medical reasons (active censoring). Most of the pregnancies occurred within 3 months after the patient had been placed on the waiting list (30 of 32, 94%). Of the 54 censored couples, 4 became pregnant.

CONCLUSIONS: On a 6-months waiting list for IVF, 13% of the couples dropped out before starting treatment. The single most important reason for dropout was (spontaneous) pregnancy. Most of these pregnancies occur within 3 months, which suggests that psychological factors such as stress relief after placed on the waiting list might be operative.

Introduction

Since the introduction of in vitro fertilization (IVF) in 1978, the indications to start an IVF treatment have been widened. In the early days the main indication for IVF was tubal occlusion, nowadays it is used for almost all categories of subfertility diagnoses.

However, in vitro fertilization can lead to physical, psychological and economical burdens, which is why counselling before starting the treatment is vital. Part of the counselling of couples before starting IVF treatment is giving them accurate information on success rates. Many studies describe pregnancy rates after starting IVF treatment; however, there is also a chance of pregnancy during the “waiting list period” (Stolwijk *et al.*, 1996; Eijkemans *et al.*, 2008; Olivius *et al.*, 2002). Patients should know what to expect in this period. Also, in an intention-to-treat analysis, dropouts of the waiting list will negatively affect the IVF centre’s cumulative pregnancy rate.

A variety of studies have described high dropout rates in IVF treatment (Stolwijk *et al.*, 1996; Olivius *et al.*, 2002; Smeenk *et al.*, 2003; De Vries *et al.*, 1999; Land *et al.*, 1997; Roest *et al.*, 1998; Verhagen *et al.*, 2008), and there is controversy in the literature about the reasons for dropout (Smeenk *et al.*, 2003; De Vries *et al.*, 1999; Roest *et al.*, 1998). Prognostic factors affecting pregnancy rates can be related to the general characteristics of patients (such as, e.g. age) or can be related to the results of previous IVF treatment. De Vries *et al.* (De Vries *et al.*, 1999) concluded that there was no selective dropout of patients with a poor prognosis. These dropouts can be actively censored (for medical reasons, as with physician-recommended drop-outs) or passively censored (as with patient-initiated dropouts). We showed that after initiation of IVF treatment the dropout rates for active censoring decreased and passive censoring increased during subsequent treatment cycles (Verhagen *et al.*, 2008). Dropouts for active or passive censoring may be considered an adverse treatment outcome because dropping out deprives a couple of additional pregnancy chances by IVF treatment and can influence the cumulative success rates per completed IVF treatment series or per IVF cycle (Verberg *et al.*, 2008).

Our study focused on couples dropping out before starting IVF, while they were still on the waiting list. We investigated the incidence of couples dropping out of the waiting list and their reasons for doing so. A better understanding of why couples leave the IVF program before starting treatment allows for tailoring patient counselling to the couple’s individual needs and at the same time provides a more realistic impression of the program’s overall cumulative pregnancy rate.

Material and methods

The indications for couples to be placed on the IVF waiting list were according to the 1998 IVF guidelines of the Dutch Society of Obstetrics and Gynaecology (NVOG-Guideline No. 9, Indications for IVF). For tubal blockage, severe endometriosis, or severe male subfertility, IVF or intracytoplasmic sperm injection (ICSI) can be offered immediately. For unexplained subfertility or mild endometriosis, IVF is offered after 3 years of subfertility, or after 2 years of subfertility when the women's age is above 36 years. Treatment such as intrauterine insemination must be offered before offering IVF. In cases of mild male subfertility or cervical factor, IVF is offered after 2 years of subfertility. For hormonal disorders, IVF is offered after 12 cycles of ovulation induction with gonadotropins. Depending on the age of the woman, IVF can be offered earlier. The Dutch guidelines are described by Lintsen *et al.* (Lintsen *et al.*, 2007).

At the time of our study, the median waiting period before initiating IVF treatment at the Maastricht University Medical Centre (MUMC) was 6 months (range: 1.2 to 19.2 months). We prospectively collected data for all 674 consecutive patients placed on the waiting list for IVF or ICSI treatment in our IVF clinic between June 2000 and July 2003. Couples using donor gametes or starting IVF for preimplantation genetic diagnosis and surgical sperm aspiration were excluded. (These indications are not mentioned in the Dutch IVF guidelines.) Patients were observed until they started their first IVF cycle or until the moment of censoring, whichever came first. If patients left the program for medical reasons, this was defined as active censoring; if patients left the program for personal reasons, this was considered passive censoring. Reasons for active censoring included failure to lose weight (body mass index >30 kg/m²), failure to gain weight (body mass index <18 kg/m²), or medical conditions such as progressive endometriosis requiring surgery, development of azoospermia while on the waiting list, hypertension requiring medical treatment, and improved semen quality no longer requiring IVF-ICSI. Dropout due to passive censoring included financial reasons, relationship problems, psychological burden, language problems, or continuation of treatment elsewhere.

A pregnancy was defined as a positive urinary pregnancy test at least 4 weeks after the last menstrual period. Follow-up data on dropouts were collected between May and December 2005, and between September and November 2008. The data were obtained by questionnaires sent to the patients, by contacting the patients' family doctor to learn whether the patient had become pregnant, and by chart review.

In counselling couples, cumulative ongoing pregnancies rates or live-birth rates are preferred (Min *et al.*, 2004). Our study analyzed the dropout on the waiting list and the reasons for dropping out, so we decided to take each couple's first pregnancy into account irrespective of its outcome.

In The Netherlands no ethics review board approval is required for retrospective chart reviews. The local ethics committee approved the collection of data for quality monitoring purposes as part of our IVF treatment protocol. Before starting their IVF treatment, all couples gave written informed consent for the use of their anonymous medical data for research purposes.

The data were analysed using Microsoft Excel (Redmond, WA) software, and 95% confidence intervals were calculated according to Wilson's method (Wilson EB, 1927).

3

Results

Between June 2000 and July 2003, 674 couples were put on the IVF waiting list. Eventually, 588 couples (87%, 95%CI 85, 90%) started their first cycle in our centre. There were 86 dropouts from the waiting list before starting IVF treatment (13%, 95%CI 10, 15%). The follow-up was nearly complete (85 of 86; 98.8%).

Analysis of patients' reasons for dropping out showed that 23 (27%; 95% CI, 19–37%) dropped out for medical reasons (active censoring), 31 (36%; 95% CI, 27–47%) due to passive censoring, and 32 because of a spontaneous pregnancy (37%; 95% CI, 28–48%). Individual reasons for active and passive censoring are summarized in Table 1.

There were 32 dropouts due to spontaneous pregnancy on the waiting list. The pregnancy rate in all 674 patients placed on the waiting list before initiating IVF was 4.8% (95% CI, 3.4–6.6%). The waiting period varied between 1.2 and 19.2 months, with a median of 6 months. Ten couples (10 of 588; 1.7%; 95% CI, 1–3%) conceived spontaneously between treatment cycles after they had started IVF (Verhagen *et al.*, 2008). Of the actively censored couples on the waiting list, two became pregnant spontaneously (8.7%), and of the passively censored couples, also two (6.5%). The cumulative pregnancy rate in the total follow-up period of 46 months was 42% (36 of 86; 95% CI, 32–52%) in all 86 dropouts from the waiting list.

TABLE 1. Reasons for dropout in 674 couples on the IVF waiting list

	Patients	95% CI	Pregnancies	95% CI
Pregnancy	32 (37%)	28-48%	32 (100%)	89-100%
Active censoring	23 (27%)	19-37%	2 (8.7%)	2-27%
medical ^a	12		0	
failure to correct overweight (BMI > 30 kg/m ²)	10		2	
failure to correct underweight (BMI < 18 kg/m ²)	1		0	
Passive censoring	31 (36%)	27-47%	2 (6.5%)	2-21%
financial	2		1	
relational	10		0	
psychological	7		0	
language problems	3		0	
treatment elsewhere	2		1	
personal	6		0	
unknown	1		-	
Total dropouts	86 (13%)	10-15%	36 (42%)	32-52%

^a *i.e. progressive endometriosis requiring surgery, developing azoospermia, severe hypertension, improved semen quality*

Indication

Dropout rates per indication for IVF treatment were similar. Most of the spontaneous pregnancies occurred in the couples with unexplained subfertility (67%). In the patients diagnosed with endometriosis no spontaneous pregnancies occurred (Table 2). There was no difference in spontaneous pregnancy rates in couples diagnosed with primary subfertility or secondary subfertility (38%, 95%CI 27, 51% vs. 48%, 95%CI 32, 65%).

Age

The mean age in all patients on the waiting list was 32.58 years (standard deviation [SD] ± 3.76 years). Most of the spontaneous pregnancies occurred in patients younger than 35 years (83%, 95%CI 68, 92%). After 35 years a decrease in spontaneous pregnancy rate occurred (Table 3).

Duration

Thirty of the 32 women (94%, 95%CI 80, 98%) who dropped out due to pregnancy became pregnant within 3 months after being placed on the waiting list (Figure 1). Eight of these

TABLE 2. Spontaneous pregnancies per indication for IVF in couples on the waiting list

Indication	Patients	Dropouts	95% CI	Pregnancies	95% CI
Unexplained	183	24 (13%)	9-19%	16 (67%)	47-82%
Endometriosis	24	3 (13%)	4-31%	0	0-56%
Male factor	355	44 (12%)	9-16%	17 (39%)	26-54%
Tubal factor	112	15 (13%)	8-21%	3 (20%)	7-45%
Total	674	86 (13%)	10-15%	36 (42%)	32-52%

TABLE 3. Spontaneous pregnancy rate per age group in women on the IVF waiting list

Age (in years)	Patients	Dropouts	95%CI	Pregnancies	95%CI
20-25	17	1 (6%)	1-27%	1 (100%)	21-100%
25-30	152	24 (16%)	11-22%	13 (54%)	35-72%
30-35	320	39 (12%)	9-16%	16 (41%)	27-57%
35-40	184	22 (12%)	8-17%	6 (27%)	13-48%
>40	1	0	0-79%	0	-
Total	674	86 (13%)	10-15%	36 (42%)	32-52%

women subsequently appeared to have been pregnant at the time they were placed on the waiting list; however, because these patients had effectively been placed on the waiting list and there was no reason to suspect pregnancy at that time, we included them in our analysis. We observed a rapid increase of pregnancy rate in the first 3 months; after that point, the increase was more gradual.

Discussion

Our study, with almost complete patient follow-up investigated the reasons for dropout in 674 women on a waiting list for IVF treatment. An important reason for dropout was pregnancy (37%), while 36% of couples dropped out for passive censoring and 27% for active censoring. The (spontaneous) pregnancy rate in patients placed on the IVF waiting list was 4.8% before the start of IVF treatment.

This is comparable with our earlier investigation (Evers *et al.*, 1998), in which 5.5% of women achieved a spontaneous pregnancy in their first 12 months on the waiting list.

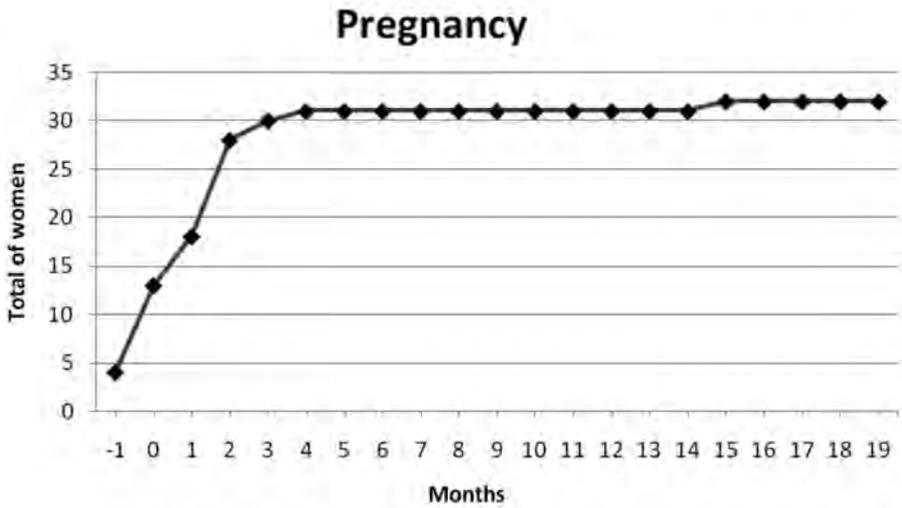


FIGURE 1. Crude cumulative pregnancy rate in couples on the IVF waiting list

Evers *et al.* (Evers *et al.*, 1998) investigated treatment-independent pregnancy rates by comparing the pregnancy rate of patients on a waiting list for IVF/ICSI with pregnancy rates of nested controls after their first IVF/ICSI cycle. The 12-month cumulative pregnancy rate of the waiting-list patients was 76 of 1391 (5.5%): 2.4% for tubal factor infertility, 5.9% for unexplained subfertility, and 6.6% for male factor infertility. More than 75% of the pregnancies in the subgroup of unexplained subfertility occurred in the first 3 months after being wait-listed.

Of the 54 couples censored for reasons other than spontaneous pregnancy in our study, four became pregnant: two (8.7%) after active censoring and two (6.5%) after passive censoring. Actively censored couples are considered to have poorer pregnancy chances than passively censored couples (1, 3). An earlier analysis by our centre of couples dropping out after the start of IVF treatment could not corroborate this assumption, showing a pregnancy rate of 25% (95% CI, 14–40%) in actively censored couples and 6% (95% CI, 2–18%) in passively censored couples (8). The suggestion was made that a reason for the higher pregnancy rate in actively censored couples might be our centre's too strict criteria for active censoring. Another reason might be that passively censored couples decide to refrain from actively pursuing conception (at least for some time) after having left the IVF program (Verhagen *et al.*, 2008). In our present study, the pregnancy rate of actively and passively censored patients was comparable with our earlier findings (Verhagen *et al.*, 2008).

The high frequency of pregnancies as a reason for dropout is an especially important factor to be considered in counselling couples before placing them on a waiting list for IVF. The pregnancies mostly occurred in couples diagnosed with unexplained subfertility and in patients younger than 35 years. Most pregnancies occurred within 3 months after the couple was placed on the waiting list. Waiting list placement may relieve the couple's stress, which may play an important psychological role in conceiving. Stress relief and its presumed positive influence on pregnancy chances were suggested by Eijkemans *et al.* (Eijkemans *et al.*, 2008) and Evers *et al.* (Evers *et al.*, 1998). Smeenk *et al.* (Smeenk *et al.*, 2003) investigated state anxiety and depression scales and concluded that passive dropout was related to psychological factors. Better counselling as well as more surveillance and support of the couples may influence dropping out for reasons related to psychological burdens.

Our aim is to minimize passive censoring, but at the same time it is important for people to put the IVF period actively to an end so that they can cope with eventual persisting childlessness (Verhagen *et al.*, 2008). Earlier reports show that it is necessary for physicians and counsellors to inform and educate the couples on success rates because after failed treatment it is a challenge for couples to make new decisions (Sydsjö *et al.*, 2005). To patients, IVF treatment might seem to offer unlimited treatment possibilities, and this may prevent couples from adopting the cognitive coping strategies they need for accepting childlessness. A well-marked treatment end point may be an important aspect of a couple's adaptation process to definitive childlessness (Verhaak *et al.*, 2007). Because ours was essentially an observational study, we did not evaluate stress and anxiety before and after putting a patient on the waiting list. Further research should be focused on the therapeutic effect of being put on the waiting list (stress relief), for example, when doing cost-benefit comparisons for IVF treatment.

Recently, Eijkemans *et al.* (Eijkemans *et al.*, 2008) reported on 5962 patients representing all patients on the IVF waiting list from January 2002 until December 2003 in the Netherlands. Initially, 6221 patients were included, and 259 (4.2%) were lost to follow-up observation. For 16% of the patients, values were missing for one or more characteristics. At the last date of follow-up evaluation, 718 patients had not become pregnant or started IVF, with a median follow-up period of 6.2 months. Their conclusion was that spontaneous pregnancy chances were low for subfertile couples, but that it is possible to identify prognostic factors (e.g., unexplained subfertility or young age) for "higher chance" couples. It can be cost-effective to postpone IVF treatment in these couples.

Our study differs from theirs in that ours is based on a smaller group and focuses on just

one clinic rather than being nationwide. However, our report is based on a longer period (from June 2000 to July 2003), and our follow-up duration was up to 46 months for the program dropouts. Furthermore, we achieved almost complete follow-up with patients who had dropped out, compared with the Eijkemans study's loss to follow-up of 4.2% and up to 16% of patients with missing values. Moreover, we not only took pregnancy rates into account but also analyzed other reasons for dropping off the waiting list.

Dropout in the IVF program compromises the cumulative pregnancy rates (Verberg *et al.*, 2008). For the patients' sake as well, dropping out should be minimized to give couples an optimal cumulative chance of conceiving. Verberg *et al.* concluded that reducing the dropout rate is necessary to improve the efficacy and cost effectiveness of IVF treatment (Verberg *et al.*, 2008). Recognition of factors that cause dropping out can help with developing interventions to prevent dropout.

An important proportion of dropouts on the waiting list for IVF are explained by the occurrence of spontaneous pregnancies. A psychological component might play a role. Critical reflection on a couple's indication for IVF before placing them on the waiting list will most likely diminish these dropouts. One might consider postponing IVF treatment in high chance groups such as patients under 35 years of age with unexplained subfertility. Better surveillance and support of couples can prevent dropping out for reasons of passive censoring. This can be done by offering patients professional support and by developing their cognitive coping strategies. Preventing dropout or preventing couples with a high chance of dropping out from being put on the waiting list may shorten the waiting list and improve patient flow. Furthermore, information about spontaneous pregnancies and patients' choices during the waiting-list period could help clinicians in counselling couples who are starting IVF treatment.

References

De Vries MJ, De Sutter P, Dhont M. Prognostic factors in patients continuing in vitro fertilization or intracytoplasmic sperm injection treatment and dropouts. *Fertil Steril* 1999; 72: 674-678.

Eijkemans MJC, Lintsen AME, Hunault CC, Bouwmans CAM, Hakkaart L, Braat DDM, *et al.* Pregnancy chances on an IVF/ICSI waiting list: a national prospective cohort study. *Hum Reprod* 2008; 23: 1627-1632.

Evers JLH, Haas HW de, Land JA, Dumoulin JCM, Dunselman GAJ. Treatment-independent pregnancy rate in patients with severe reproductive disorders. *Hum Reprod* 1998; 13: 1206-1209.

Land JA, Courtar DA, Evers JLH. Patient dropout in an assisted reproductive technology program: implications for pregnancy rates. *Fertil Steril* 1997; 68: 278-281.

Lintsen AME, Eijkemans MJC, Hunault CC, Bouwmans CAM, Hakkaart L, Habbema JDF, *et al.* Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Hum Reprod* 2007; 22: 2455-62.

Min JK, Breheny SA, MacLachlan V, Healy DL. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. *Hum Reprod* 2004; 19: 3-7.

Olivius K, Friden B, Lundin K, Bergh C. Cumulative probability of live birth after three in vitro fertilization/intracytoplasmic sperm injection cycles. *Fertil Steril* 2002; 77: 505-510.

Roest J, Van Heusden AM, Zeilmaker GH, Verhoeff A. Cumulative pregnancy rates and selective drop-out of patients in in-vitro fertilization treatment. *Hum Reprod* 1998; 13: 339-341.

Smeenk JMJ, Verhaak CM, Stolwijk AM, Kremer JAM, Braat DDM. Reasons for dropout in an in vitro fertilization/intracytoplasmic sperm injection program. *Fertil Steril* 2003; 81: 262-268.

Stolwijk AM, Hamilton CJCM, Hollanders JMG, Bastiaans LA, Zielhuis GA. A more realistic approach to the cumulative pregnancy rate after in-vitro fertilization. *Hum Reprod* 1996; 11: 660-663.

Sydsjö G, Ekholm K, Wadsby M, Kjellberg S, Sydsjö A. Relationships in couples after failed IVF treatment: a prospective follow-up study. *Hum Reprod* 2005; 20: 1952-1957.

Verberg MFG, Eijkemans MJC, Heijnen EMEW, Broekmans FJ, De Klerk C, Fauser BCJM, *et al.* Why do couples drop-out from IVF treatment? A prospective cohort study. *Hum Reprod* 2008; 23: 2050-2055.

Verhaak CM, Smeenk JMJ, Evers AWM, Kremer JAM, Kraaimaat FW, Braat DDM. Women's emotional adjustment to IVF: a systematic review of 25 years of research. *Hum Reprod Update* 2007; 13: 27-36.

Verhagen TEM, Dumoulin JCM, Evers JLH, Land JA. What is the most accurate estimate of pregnancy rates in IVF dropouts? *Hum Reprod* 2008; 23: 1793-1799.

Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927; 22: 209-212.

Chapter 4

What is the most accurate estimate of pregnancy rates in IVF dropouts?

Tamara E.M. Verhagen

John C.M. Dumoulin

Johannes L.H. Evers

Jolande A. Land

Human Reproduction 2008;23:1793-1799

Abstract

BACKGROUND: Dropouts in IVF-programmes affect cumulative pregnancy rates (CPRs), but it is unknown what the impact of loss to follow-up is.

METHODS: Data were obtained from 588 couples starting IVF treatment ('as treated group'). Cycle-based and real-time-based CPRs were calculated using three assumptions for dropouts: dropouts having no probability of pregnancy, dropouts having the same probability of pregnancy as those continuing treatment and dropouts stopping because of medical reasons having no chance of pregnancy and those stopping because of other reasons having the same probability of pregnancy as those continuing treatment. CPRs obtained in the 'as treated group' were compared to CPRs calculated using the data set including the follow-up data of the dropouts ('completed group').

RESULTS: In 1.7% of couples, no follow-up could be obtained. The cycle-based CPR after three IVF-cycles ranged from 63-71% in the 'as treated group' and was 65% in the 'completed group'. The real-time-based CPR after nine months ranged from 54-59% in the 'as treated group' and was 55% in the 'completed group'. The PR in dropouts was 14% (95% Confidence Interval 8.22%).

CONCLUSIONS: In IVF programmes, outcome data of dropouts remain unknown, and CPRs should be calculated by assuming dropouts to have a PR between no probability and the same probability as those who continue treatment. Our study shows that the most accurate estimate for the PR in dropouts is 14%.

Introduction

Accurate information on success rates is a prerequisite for adequate counselling of couples before starting their IVF treatment. There are several methods of estimating the effectiveness of treatment. The method most often used is the life table analysis based on cumulative pregnancy rates (CPRs). This method for reporting on infertility treatment outcome was introduced > 30 years ago (Lamb and Cruz, 1972). Life table analysis estimates the percentage of patients who will conceive during a particular period of follow-up, and it assumes that results in patients who drop out do not differ substantially from those who continue treatment.

It has been shown in a computer simulation program that life table analysis tends to overestimate pregnancy rates (PRs) (Doody, 1993), in particular in a population with low monthly fecundity rates and low 'cure' rates. These couples might be expected to be more easily discouraged from continuing treatment and to have higher dropout rates (Doody, 1993). Other studies showed that making different assumptions concerning the prognosis of couples discontinuing IVF treatment affected the CPR significantly (Stolwijk *et al.*, 1996, Olivius *et al.*, 2002). If one assumes that the pregnancy chance of dropouts is similar to the one in those who continue treatment, the CPR will be overestimated. On the other hand, if dropouts are assumed not to have any probability of getting pregnant, the CPR will be underestimated. It has been suggested that the most realistic estimate of the true CPR might be obtained by assuming that dropouts because of medical reasons (i.e. poor prognosis and active censoring) have no chance of pregnancy, and that those who stop because of other, more personal reasons (passive censoring) have the same chance of pregnancy as those who continue treatment (Stolwijk *et al.*, 1996, Olivius *et al.*, 2002).

Stolwijk *et al.* calculated the CPR after five completed IVF cycles for these different assumptions in dropouts. The CPR was 51% when dropouts were assumed to have the same probability of pregnancy as those continuing treatment, 37% when dropouts were assumed to have no probability of pregnancy at all and 41% when dropouts due to passive censoring were assumed to have the same probability of pregnancy as those who continued treatment whereas actively censored dropouts had no chance of pregnancy (Stolwijk *et al.*, 1996). Olivius *et al.* calculated the CPR after three completed cycles of IVF using the same assumptions. The CPR was 73% assuming all dropouts had the same probability of pregnancy as those continuing treatment, 65% assuming dropouts had no chance of pregnancy and 66% assuming dropouts due to passive censoring

had the same chance of pregnancy as those who continued treatment whereas actively censored dropouts had no chance of pregnancy (Olivius *et al.*, 2002). Both studies lacked information on the true PR in the dropouts.

In a recent review, several pitfalls of life table analysis have been summarized (Daya, 2005). Daya proposed several solutions for generating more realistic outcome data. One strategy to correct for the overestimation in CPR is to use the cycle based CPR assuming that all dropouts have a 0% PR in subsequent cycles. This approach will underestimate the CPR, but this will encourage clinics to be more realistic in counselling couples about prognosis of IVF treatment. Another strategy is the construction of a real-time-based CPR. After agreement on the average length of time a couple needs to complete one cycle of IVF treatment, involving fresh and frozen embryo transfers, the CPR can be calculated for any length of time. The outcome is summarized as the CPR after a specified period of time. However, in this strategy there is no solution for the dropouts due to active or passive censoring. This can be solved by assuming all dropouts (whatever their reason for dropout) to have a 0% PR for the remaining study period (Daya, 2005). Daya recommends the real-time-based CPR for calculating treatment outcome because it focuses on the couple as the unit of analysis and it satisfies the time scale requirement of the life table method. The problem caused by the unknown outcome in the dropouts can only be corrected for by having access to dropout follow-up data, both with and without therapy.

In the present study, dropouts were traced and their pregnancies after dropping out of the IVF programme calculated, in order to arrive at the best estimate of the overall CPR in couples undergoing IVF. Different methods to calculate CPR were compared and we deduced the best method for reporting IVF outcome in everyday practice, in which outcome in dropouts is usually unknown.

Material and methods

Data from all 588 couples starting their first IVF or ICSI-treatment in our IVF clinic between June 2000 and July 2003 were collected retrospectively. Couples starting IVF for preimplantation genetic diagnosis and surgical sperm aspiration or those using donor gametes were excluded. In the Netherlands, health retrieval reimburse three IVF cycles, and therefore a maximum of three completed cycles was included in the study. A cycle was considered completed when ovum retrieval was performed. Transfers of cryopreserved

embryos were considered to be part of the fresh cycles from which they originated.

Patients were observed until either a pregnancy was achieved, until they withdrew from the programme before a pregnancy was established or until three IVF cycles had been completed. A pregnancy was defined as a positive urinary pregnancy test 14 days after embryo transfer. All pregnancies achieved during the period a couple underwent IVF were considered treatment-dependent pregnancies.

Patients withdrew from the programme when further treatment was denied for medical reasons (active censoring), or when the couple abandoned further treatment because of personal reasons (passive censoring). Reasons for active censoring were poor response, poor fertilisation, poor response with poor fertilisation, overweight with BMI > 30 kg/m², hypertension or improved semen quality not requiring ICSI any more. Reasons for dropout due to passive censoring were relational problems, additional health problems, psychological burden, physical burden, both psychological and physical burden, and continuation of treatment elsewhere.

Patients who did not inform staff members about their decision to abandon or postpone treatment and who failed to return to follow-up appointments > 6 months were considered lost to follow-up.

The computer-based IVF database and the patients' medical records were used for data collection. Follow-up data of dropouts were collected between May and December 2005, and were obtained by questionnaires sent to the patients and by contacting the patients' general practitioner to learn whether the patient had become pregnant. All pregnancies in the dropouts were considered treatment-independent pregnancies, including pregnancies after treatment in other clinics.

For the calculations, two different datasets were constructed: (1) the outcome data of the couples that underwent IVF treatment, defined as the 'as treated group' and (2) the data from the treated group supplemented with the achieved follow-up data from the dropouts, defined as the 'completed group'.

The PR and CPR were calculated for the 'as treated group' and the 'completed group' using the cycle-based and real-time-based methods respectively, as proposed by Daya (Daya, 2005). For the cycle-based CPR, a maximum of three treatment cycles was used. For the real-time-based CPR the length of time for one completed cycle (including fresh

and frozen embryo transfers) was arbitrarily set at three months, since in our program 85% of couples completed one IVF cycle within this period.

The data set of the 'as treated group' was used to estimate the cycle-based and real-time-based PR and CPR with the assumptions usually made for dropouts (Stolwijk *et al.*, 1996, Olivius *et al.*, 2002): (1) assuming dropouts had no chance of pregnancy (in the remainder of the manuscript this is abbreviated as 'dropouts=0%'), (2) assuming dropouts had the same chance of pregnancy as those who continued treatment ('dropouts=same') and (3) assuming dropouts due to active censoring had no chance of pregnancy and those dropping out due to passive censoring had the same chance of pregnancy as those who continued treatment ('dropouts=differentiated').

The data set of the 'completed group' was used to calculate the effective cycle-based and real-time-based PR and CPR. The CPR's from the treated group with the three different assumptions for PR in the dropouts were compared to the 'true' CPR from the completed group.

In the Netherlands, for retrospective chart review, no ethical review board approval is required. Before starting their IVF treatment, all couples had given written informed consent for the use of their anonymised medical data for research purposes.

The data were analysed using Microsoft Excel software. 95% Confidence Intervals (CI) were calculated according to Wilson's method (Wilson, 1929).

Results

Between June 2000 and July 2003 588 couples started their first cycle in our IVF centre. Together they completed 1024 cycles. The characteristics for the total group, the couples completing IVF treatment and those dropping out are listed in Table 1. There were no significant differences in composition between the two groups of couples.

Of the 588 couples starting IVF, 480 (82%) completed treatment (i.e. became pregnant or completed three treatment cycles) and 108 (18%) dropped out. Before the first cycle was completed 26 couples (24% of all dropouts) had stopped: 20 because of poor response (active censoring) and 6 for personal reasons (passive censoring). After the first cycle another 47 couples (44% of all dropouts) withdrew from the programme (21 actively and 26 passively censored). After the second cycle 35 couples dropped out (32%) (10 actively and 25 passively censored). Reasons for active censoring were poor response (n=29), poor fertilisation (n=10), poor response with poor fertilisation (n=6), overweight with BMI > 30 kg/m² (n=3), hypertension (n=1) and improved semen quality not requiring

TABLE 1. Characteristics of 588 couples (mean \pm SD) who started IVF treatment between June 2000- July 2003, and either completed treatment (up to three cycles or until pregnancy) or dropped out.

	Completed IVF (n = 480)		Dropped out (n = 108)		p
Age of female (years)	32.9	\pm 3.6	33.8	\pm 4.1	NS
Duration of subfertility (years)	3.0	\pm 2.2	3.5	\pm 2.4	NS
Primary subfertility	346	(72.1%)	68	(63%)	NS
Cause of subfertility:					
Male factor	256	(53.3%)	55	(50.9%)	NS
Anovulation	2	(0.4%)	1	(0.9%)	NS
Endometriosis	17	(3.5%)	4	(3.7%)	NS
Tubal factor	75	(15.6%)	22	(20.4%)	NS
Unexplained	130	(27.1%)	26	(24.1%)	NS
IVF treatment	219	(45.6%)	51	(47.2%)	NS
ICSI treatment	261	(54.4%)	57	(52.8%)	NS

NS = Not Significant

4

TABLE 2. Pregnancies after IVF treatment (“as treated group”) and pregnancies in dropouts with known follow-up.

	IVF treatment		Dropout ^a	
Couples	588		108	
Cycles	1024		NA	
Pregnancy	368		15 ^a	
Live birth	278	(75.5%)	13	(86.7%)
Miscarriage, ectopic	76	(20.6%)	1	(6.7%)
Fetal death	4	(1.1%)	0	
Unknown outcome	10	(2.7%)	1	(6.7%)

NA = not available

^a 12 spontaneous pregnancies and three pregnancies after treatment in other hospitals

ICSI any more (n=2). Reasons for dropout due to passive censoring were relational problems (n=6), additional health problems (n=3), psychological burden (n=8), physical burden (n= 4), both psychological and physical burden (n=18), continuation of treatment elsewhere (n=6), and unknown reasons (n=12).

There were 368 pregnancies in the treated couples (‘as treated group’), 325 after fresh embryo transfers and 43 after frozen embryo transfers. These pregnancies resulted in 278 live births (76%), 251 after fresh transfers (77%) and 27 after frozen transfers (63%). Details are given in Table 2.

TABLE 3. Cycle-based Pregnancy Rate (PR) and Cumulative Pregnancy Rate (CPR) using the dataset from the “completed group”, i.e. treated patients (IVF) and dropouts.

Cycle No.	Couples (n)	Pregnancies (n)			PR (%)	CPR (%)	95% CI
		Total	IVF	dropouts			
1	588	226	224	2	38	38	34,42
2	362	115	106	9	32	58	54,62
3	247	42	38	4	17	65	61,69

In 98 of the 108 dropouts (91%), follow-up data were secured by sending questionnaires or contacting the general practitioners. Only ten couples (1.7% of all couples who entered the programme) were lost to follow-up completely, of them three were actively censored and seven passively. There were 15 pregnancies in the 98 dropouts (15%): three patients conceived after fertility treatment at another clinic and 12 patients conceived spontaneously, resulting in 13 live births. Of the 51 actively censored couples (follow-up available in 48) 12 did eventually become pregnant (25%) and of the 57 passively censored couples (follow-up in 50), three conceived (6%).

Calculation of CPR using different strategies

1. Cycle-based CPR. The cycle-based CPR and 95% CI for the ‘completed group’ was 65% (61,69%) (Table 3). Table 4 shows the cycle-based CPR using the dataset from the ‘as treated group’. For these calculations three different assumptions were made for the dropouts: ‘dropouts=0%’, ‘dropouts=same’ and ‘dropouts=differentiated’. The cycle-based CPR and 95% CI were 63% (59,67%), 71% (67,74%) and 66% (62,70%) respectively. Figure 1 shows the different cycle-based CPR’s and illustrates that the CPR in

TABLE 4. Cycle-based Pregnancy Rate (PR) and Cumulative Pregnancy Rate (CPR) using the dataset from the “as treated group” and the different assumptions made for dropouts.

Cycle no.	Pregnancies (n)	Dropouts		Dropouts = 0% ^a			
		Active	Passive	Couples (n)	PR (%)	CPR (%)	95% CI
1	224	41	32	588	38	38	34,42
2	106	10	25	364	29	56	52,60
3	38	-	-	258	15	63	59,67

^a assuming dropouts had no chance of pregnancy (“dropouts = 0%”)

^b assuming dropouts had the same chance of pregnancy as those who continued treatment (“dropouts = same”)

^c assuming dropouts due to active censoring had no chance of pregnancy and those dropping out due to passive censoring had the same chance of pregnancy as those who continued treatment (“dropouts = differentiated”)

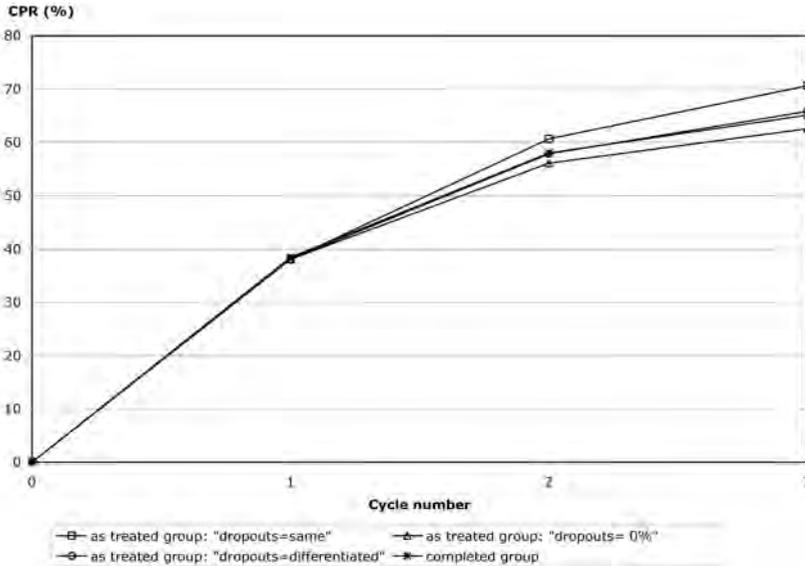


FIGURE 1. Cycle-based Cumulative Pregnancy Rates (CPR) for the “as treated group” with the different assumptions for the dropouts and the “completed group”.

the ‘completed group’ corresponds best with the CPR calculated by applying the ‘dropouts=differentiated’ assumption.

2. Real-time-based CPR. This strategy for calculating the CPR is based on the time to pregnancy. The time needed to complete one cycle including fresh and frozen embryo transfers was set at three months, and therefore the time needed for three cycles was nine months. In Table 5 the results for the ‘completed group’ are given. In one couple who stopped treatment the time to pregnancy was not known. The CPR and 95% CI is

	Dropouts = same ^b				Dropouts = differentiated ^c			
	Couples (n)	PR (%)	CPR (%)	95% CI	Couples (n)	PR (%)	CPR (%)	95% CI
	588	38	38	34,42	588	38	38	34,42
	291	36	61	57,65	332	32	58	54,62
	150	25	71	67,74	201	19	66	62,70

TABLE 5. Real-time based Pregnancy Rate (PR) and Cumulative Pregnancy Rate (CPR) per period of three months for the “completed group”, i.e. treated patients (IVF) and dropouts (DO).

Period (months)	Couples (n)	Pregnancies IVF (n)	Pregnancies dropouts (n)	PR (%)	CPR (%)	95% CI
< 3	588	207	2	36	36	32,40
3-6	380	75	0	20	48	44,52
6-9	302	38	2	13	55	51,59
9-12	229	21	1	10	59	55,63
12-15	184	12	2	8	63	59,67
15-18	152	11	1	8	66	62,70
18-21	122	1	1	2	66	62,70
21-24	112	1	2	3	67	63,71
> 24	105	2	3	5	69	65,73

55% (51,59%) after nine months for the ‘completed group’. In Table 6 the CPR for the ‘as treated group’ is given, using the three different assumptions concerning the dropouts (‘dropouts=0%’, ‘dropouts=same’ and ‘dropouts=differentiated’). These expected real-time CPR’s and 95% CI were 54% (50,58%), 59% (55,63%) and 56% (52,60%) respectively after nine months.

TABLE 6. Cycle-based Pregnancy Rate (PR) and Cumulative Pregnancy Rate (CPR) using the dataset from the “as treated group” and the different assumptions made for dropouts.

Period (months)	Pregnancies (n)	Dropouts (n)		Dropouts = 0% ^a			
		Active	Passive	Couples (n)	PR (%)	CPR (%)	95% CI
< 3	207	27	24	588	35	35	31,39
3-6	75	15	22	381	20	48	44,52
6-9	38	5	5	303	13	54	50,58
9-12	21	0	4	232	9	59	55,63
12-15	12	2	0	188	6	61	57,65
15-18	11	1	2	158	7	64	60,68
18-21	1	1	0	129	1	64	60,68
21-24	1	0	0	120	1	65	58,66
> 24	2	0	0	115	2	65	58,66

^a assuming dropouts had no chance of pregnancy (“dropouts = 0%”)

^b assuming dropouts had the same chance of pregnancy as those who continued treatment (“dropouts = same”)

^c assuming dropouts due to active censoring had no chance of pregnancy and those dropping out due to passive censoring had the same chance of pregnancy as those who continued treatment (“dropouts = differentiated”)

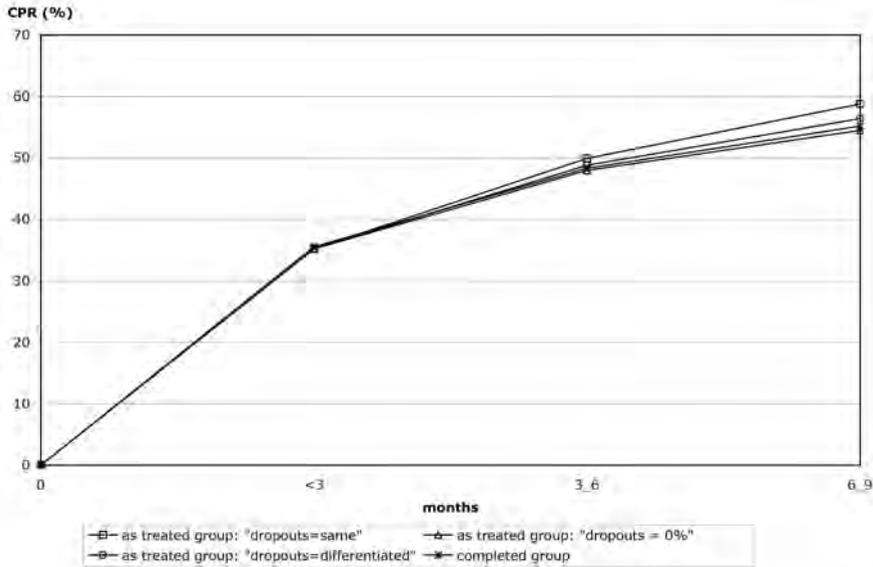


FIGURE 2. Real-time based Cumulative Pregnancy Rate (CPR) using the dataset of the “as treated group” and the different assumptions made for the dropouts.



Figure 2 shows the real-time-based CPR for the different subgroups for nine months. The CPR in the ‘completed group’ corresponds best with the CPR calculated by applying the ‘dropouts=differentiated’ assumption.

	Dropouts = same ^b				Dropouts = differentiated ^c			
	Couples (n)	PR (%)	CPR (%)	95% CI	Couples (n)	PR (%)	CPR (%)	95% CI
	588	35	35	31,39	588	35	35	31,39
	330	23	50	46,54	357	21	49	45,53
	215	18	59	55,63	257	15	56	52,60
	134	16	65	61,69	181	12	61	57,65
	86	14	70	66,74	133	9	65	61,69
	54	20	76	72,79	103	11	69	65,73
	22	5	77	73,80	72	1	69	65,73
	12	8	79	76,82	63	2	70	66,74
	7	29	85	82,88	58	3	71	67,75

Discussion

The cycle-based CPR and real-time CPR were calculated for the 'as treated group' and the 'completed group', using data from 588 couples. The 'completed group' included the follow-up data from dropouts. In most studies performed a substantial number of follow-up data in dropouts is missing. In our study follow-up was complete in 98.3% of all couples starting treatment, and in 98 of 108 couples who dropped out (91%) follow-up data could be secured. In this 'completed group' a 65% cycle-based CPR after three cycles and a 55% real-time-based CPR after nine months was calculated. For the 'as treated group' CPR was calculated using different assumptions for the dropouts. For the cycle-based CPR the 'dropouts=differentiated' assumption was found to correspond to the cycle-based CPR in the 'completed group' (65% and 66% respectively). For the real-time-based CPR, after nine months of treatment the CPR in the 'completed group' (55%) was comparable to the CPR for 'dropouts=0%' (54%) and the 'dropout=differentiated' (56%) in the 'as treated group'. These findings indicate that the impact of dropouts is smaller on real-time-based CPR's as compared to cycle-based CPR's.

In most of the studies presenting CPR's different assumptions for dropouts are made. Incorporating the reason for dropout (in this study 'dropouts =differentiated') is assumed to give the most realistic estimation of the CPR (Stolwijk *et al.*, 1996, Olivius *et al.*, 2002), as actively censored patients are considered to have poor pregnancy chances and passively censored patients to have a better prognosis. Although our study confirmed that the strategy in which the reason for dropping out was taken into account gave the most realistic CPR, the generally held assumption concerning PR in actively and passively censored patients could not be corroborated. There were 12 pregnancies in 48 actively censored couples (25%, 95% CI 14,40%) and only three pregnancies in 50 passively censored couples (6%, 95% CI 2,18%). There were three actively censored patients lost to follow-up and seven passively censored ones. We have performed a sensitivity analysis and assumed, according to a worst-case scenario, that none of the dropouts lost to follow-up did conceive. CPR's were also calculated after extrapolating the pregnancies in the known dropouts: 12 pregnancies in actively censored couples (follow-up in 48 of 51 couples) and three in passively censored couples (follow-up in 50 of 57 couples). Extrapolation of these data meant one more pregnancy in the actively censored couples. This did not affect the CPR's. We did not perform a best-case scenario since the findings in the known dropouts indicated a high pregnancy rate in the dropouts lost to follow-up to be unrealistic, but even if all seven passively censored patients would have conceived,

the resulting PR of 20% in the passively censored patients would not have exceeded the PR of 25% in the actively censored group. Reasons for the PR in actively censored patients to be higher than in the passively censored group in our study might be due to (too) strict criteria for active censoring, or to passively censored couples refraining from trying to conceive after censoring. This study offers support for the contention that the differentiated method offers the best estimate of the eventual CPR, both cycle and real-time-based. From the group with complete follow-up in our study we calculated the PR in dropouts to be 14% (95% CI 8,22%).

In our study a pregnancy was defined as a positive urinary pregnancy test 14 days after the embryo transfer. We are aware that for counselling couples cumulative ongoing pregnancies rates or live birth rates are preferred (Min *et al.*, 2003). As the aim of our study was to evaluate the effect of different datasets and models on CPR, it was decided to take the first pregnancy in a couple into account irrespective of its outcome. In the 368 pregnancies in the 'as treated group' the live birth rate was 76% (Table 2), which is in accordance with other studies (Simon *et al.*, 1999; Tummers *et al.*, 2003).

In our study no further follow-up was obtained in 112 couples who completed three IVF cycles without becoming pregnant or dropping out. After three IVF cycles most couples stop treatment in the Netherlands because additional cycles are not reimbursed, and only a small percentage of patients will continue treatment at their own expense.

In our previous study in 1997 a dropout rate of 26% was found after the first cycle, and 32% of these dropouts were due to active censoring. After the second cycle the dropout rate was 29%, of which 29% was due to active censoring (Land *et al.*, 1997). These dropout rates are in accordance with other studies published so far, reporting rates between 22% and 40% after the first IVF cycle (Roest *et al.*, 1998, de Vries *et al.*, 1999, Olivius *et al.*, 2002, Schroder *et al.*, 2004). In the present study the dropout rates were lower, namely 12% after the first and 12% after the second cycle, and the percentage of actively censored patients was 56 and 29, respectively. The decreasing percentage of couples dropping out from our IVF programme between the two study periods might be due to better patient counselling, surveillance of their progress and more structural support for couples between treatment cycles. This might have decreased dropping out for reasons related to the psychological and physical burden of the IVF treatment. We strive for minimizing passive censoring, since our couples have taught us the importance of actively putting a period to an end in order to be able to cope with their eventual persisting childlessness.

In conclusion: for counselling couples who start treatment and for reporting on IVF-results, CPR is to be preferred instead of PR per cycle. In countries in which the number of treatment cycles is fixed, cycle-based CPR is the best outcome parameter for reports on IVF. Overestimation or underestimation of CPR due to invalid assumptions of PR in dropouts should be minimized and CPR should be calculated by assuming dropouts to have a PR between no probability of pregnancy at all and the same probability of pregnancy as those who continue treatment. Our study shows that the most accurate estimate of the PR in dropout is 14% (95% CI 8,22%).

References

- Daya S (2005) Life table (survival) analysis to generate cumulative pregnancy rates in assisted reproduction: are we overestimating our success rates? *Hum Reprod* 20,1135-1143.
- Doody MC (1993) Drop-out behaviour and fertility table analysis of pregnancy rates. *Hum Reprod* 8,886-889.
- Hull MGR (1994) Infertility Services. In Maresh M (ed.) *Audit in Obstetrics and Gynaecology*. Blackwell, London, pp.197-245.
- Lamb EJ and Cruz AL (1972) Data collection and analysis in an infertility practice. *Fertil Steril* 23,310-319.
- Land JA, Courtar DA and Evers JLH (1997) Patient dropout in an assisted reproductive technology program: implications for pregnancy rates. *Fertil Steril* 68,278-281.
- Min JK, Breheny SA, MacLachlan V, Healy DL (2004) What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. *Hum Reprod* 19:3-7.
- Olivius K, Friden B, Lundin K and Bergh C (2002) Cumulative probability of live birth after three in vitro fertilization/ intracytoplasmic sperm injection cycles. *Fertil Steril* 77,505-510.
- Roest J, van Heusden AM, Zeilmaker GH, Verhoeff A (1998) Cumulative pregnancy rates and selective drop-out of patients in in-vitro fertilization treatment. *Hum Reprod* 13:339-341.
- Schroder AK, Katalinic A, Diedrich K and Ludwig M (2004) Cumulative pregnancy rates and drop-out rates in a German IVF programme: 4102 cycles in 2130 patients. *Reprod Biomed Online* 8(5),600-606.
- Simon C, Landeras J, Zuzuarregui JL, Martin JC, Remohi J, Pellicer A (1999) Early pregnancy losses in in vitro fertilization and oocyte donation. *Fertil Steril* 72:1061-1065.
- Sharma V, Allgar V and Rajkhowa M (2002) Factors influencing the cumulative conception rate and discontinuation of in vitro fertilization treatment for infertility. *Fertil Steril* 78,40-46.
- Stolwijk AM, Hamilton CJCM, Hollanders JMG, Bastiaans LA and Zielhuis GA (1996) A more realistic approach to the cumulative pregnancy rate after in-vitro fertilization. *Hum Reprod* 11,660-663.
- Tummers P, De Sutter P, Dhont M (2003) Risk of spontaneous abortion in singleton and twin pregnancies after IVF/ICSI. *Hum Reprod* 18:1720-1723
- Vries MJ de, De Sutter P, Dhont M (1999) Prognostic factors in patients continuing in vitro fertilization or intracytoplasmic sperm injection treatment and dropouts. *Fertil Steril* 72,674-678.
- Wilson EB (1929) Probable Inference, the Law of Succession, and Statistical Inference. *J Am Stat Assoc* 22,209-212.

Chapter 5

Impact of dropout during fertility treatment on the estimation of pregnancy rates in assisted reproduction

Tamara E.M. Verhagen

Inge M. Custers

Pieterneel Steures

Ben Willem J. Mol

Johannes L.H. Evers

submitted

Abstract

It has been suggested that pregnancy rates reported in ART studies are inflated due to high dropout rates, but empirical data are lacking. A literature search was performed for studies reporting on patients undergoing multiple cycles of IUI or IVF/ICSI. Studies were included if they allowed extraction of dropout and pregnancy rate per cycle. For each cycle dropout and pregnancy rates were calculated. Data on IUI and IVF were analysed separately. We used Pearson correlation coefficient to evaluate whether dropout rate was associated with pregnancy rate. Overall, we analysed data of 17,555 couples, who underwent 52,003 IUI cycles and of 45,934 couples, who underwent 99,345 IVF cycles. For IUI, pregnancy rates ranged from 0% to 50% per cycle (mean 8%; 95% CI 7, 9) and dropout rates ranged from 4% to 85% per cycle (mean 30%; 95% CI 27, 33). For IVF, pregnancy rates ranged from 7% to 36% per cycle (mean 18%; 95% CI 16, 19), and dropout rates ranged from 7% to 69% per cycle (mean 36%; 95% CI 32, 41). For IUI and IVF we found no significant relation between pregnancy rate and dropout rate in the previous cycle. Overestimation of pregnancy rates in ART due to dropout is limited.

Introduction

Intrauterine insemination (IUI) and in vitro fertilization (IVF) are commonly used procedures in the treatment of a wide range of fertility problems. The effectiveness of IUI and IVF is usually expressed as pregnancy rate per treatment cycle or cumulative pregnancy rates after several cycles. Dropout of couples from assisted reproductive technology (ART) programmes is very common and occurs at any treatment stage (Olivius *et al.*, 2004; Brandes *et al.*, 2009; Eisenberg *et al.*, 2010). Dropout tends to increase during consecutive treatment cycles and can be as high as 75% per cycle (Dickey *et al.*, 2002; Wainer *et al.*, 2004; Byrd *et al.*, 1987; Custers *et al.*, 2008; Fahri *et al.*, 2010). Patients' dropout can be divided in dropout due to voluntarily discontinuation of treatment and dropout due to physicians' recommendation because of poor prognostic factors. These types of dropout are also known as passively and actively censoring of patients respectively (Land *et al.*, 1997). In calculations of cumulative pregnancy rates, it is assumed that couples that discontinue treatment would have had the same probability of conception as those who continue with multiple attempts. The assumption that dropout is not related to the outcome probably is satisfied for couples who discontinue ART for reasons such as moving to another city, who suffer a relationship breakdown or who experience practical problems such as travelling distance to the clinic and difficulties to take time off from work. However, if patients' dropout is related to the outcome of a particular treatment, the estimate of (cumulative) pregnancy rates will be affected (Land *et al.*, 1997; Daya 2005). It has been suggested that actively censored couples and dropouts for psychological reasons in ART have a substantially worse fertility prognosis than uncensored patients, who continue treatment (Sharma *et al.*, 2002; Smeenk *et al.*, 2004; Haan *et al.*, 1991). This phenomenon may well inflate estimates of pregnancy rates in consecutive treatment cycles (Daya, 2005).

We hypothesized that a high dropout rate is associated with a high pregnancy rate in the next cycle. Since empirical data on this issue are lacking, the aim of this study was to explore whether empirical data can confirm or refute a relationship between pregnancy rates and patient dropout rates in consecutive ART treatment cycles.

Materials and methods

A PubMed search of studies reporting on pregnancy rates after IUI and IVF in multiple consecutive treatment cycles from 1980 to 2012 was performed. Search terms for IUI were: (“Pregnancy Rate”) AND (“Insemination, Artificial, Homologous”, OR “Insemination Artificial, Heterologous”) AND (“cycle” OR “cycles”). Search terms for IVF were: (“Pregnancy Rate”) AND (“Fertilization in Vitro”) AND (“therapy” OR “therapeutics”) AND (“cycle” OR “cycles”). Limits were set on human and English. Only studies that reported on pregnancy outcome per treatment cycle were included. The following characteristics of each study were recorded: (1) data collection (prospective versus retrospective), (2) treatment given (IUI with/ without controlled ovarian hyperstimulation; IVF and/or ICSI), (3) indication for treatment and (4) the definition of pregnancy (positive urinary test, gestational sac on ultrasound, cardiac activity on ultrasound at 9-12 weeks, or live birth).

For each study pregnancy rates were calculated per treatment cycle. Since dropout can only be calculated after the first cycle, cycles were included from the second cycle onwards. Dropout numbers were calculated by subtracting the number of patients starting in the

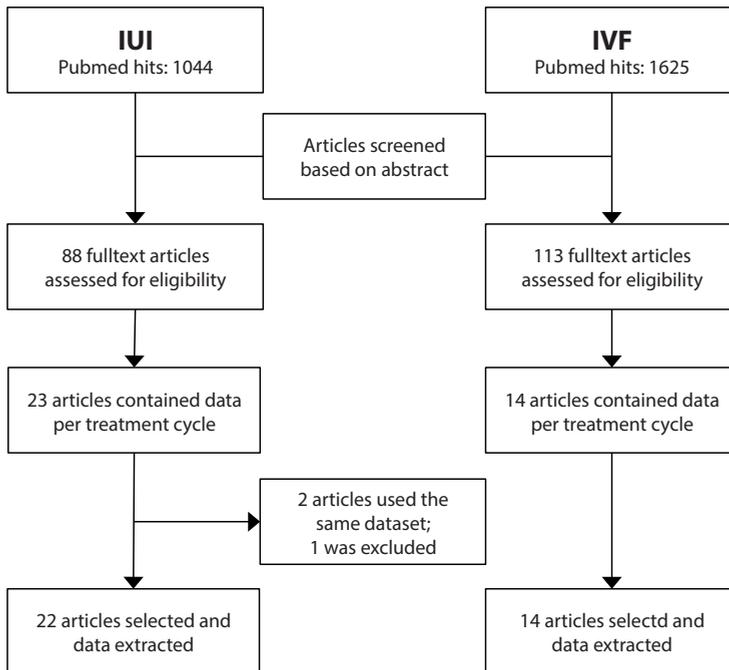


FIGURE 1. Flow diagram of study selection

current cycle from patients who did not get pregnant in the previous cycle. The dropout rate was calculated by dividing the number of dropouts by the number of patients who did not get pregnant in the previous cycle. For IUI dropout rates and pregnancy rates were calculated up to the ninth treatment cycle, or less if not all variables were reported until the ninth cycle. For IVF, the same was done until the sixth treatment cycle, as far as possible.

We calculated mean dropout rates in the previous cycle and pregnancy rates in the current cycle and 95% confidence intervals. The relation between pregnancy rates and treatment cycle in consecutive treatment cycles was analysed with the Pearson correlation coefficient, with pregnancy rate being the dependent variable and cycle number the independent variable. The effect of dropout rates on pregnancy rates was analysed using the Pearson correlation coefficient, with dropout rate being the independent variable and pregnancy rate the dependent variable. This analysis was done separately for the effect of dropout after the first cycle on the pregnancy rate in the second cycle. We assumed that the information after one treatment attempt provides the most reliable data on dropout rate in relation to the pregnancy rate in the next cycle. The analyses were stratified for IUI and IVF. Data collection and statistical analysis were performed using Microsoft Excel 2008.

Results

The search on IUI resulted in 23 studies (Byrd *et al.*, 1987; Lalich *et al.*, 1988; Dodson and Haney 1991; Plosker *et al.*, 1994; Campana *et al.*, 1996; Berg *et al.*, 1997; Hendin *et al.*, 2000; Khalil *et al.*, 2001; Dickey *et al.*, 2002; Lee *et al.*, 2002; Steures *et al.*, 2004; Wainer *et al.*, 2004; Ibérico *et al.*, 2004; Dickey *et al.*, 2005; van Rumste *et al.*, 2006; Dankert *et al.*, 2007; Custers *et al.*, 2008; Vitthala *et al.*, 2008; de Brucker *et al.*, 2009; Fahri and Orvieto, 2010; Merviel *et al.*, 2010; Dorjpurev *et al.*, 2011; Groeneveld *et al.*, 2011). The results of the IUI cycles reported in the study of Steures *et al.*, 2004 were part of the study published in 2008 from the same group (Custers *et al.*, 2008). The later study included data from a longer period and hence was used for the IUI results on pregnancy and dropout rates, so data from 22 studies reporting on IUI were available (Figure 1).

The search on IVF resulted in 14 studies reporting that fulfilled the inclusion criteria (Guzick *et al.*, 1986; Sharma *et al.*, 1988; Haan *et al.*, 1991; Tan *et al.*, 1992; French National IVF Registry 1993; Check *et al.*, 1994; Roest *et al.*, 1998; Olivius *et al.*, 2002;

TABLE 1. Study characteristics of selected studies on IUI and IVF

Authors	Year	Retrospective / Prospective	Treatment	Diagnosis	Couples	Number of cycles in original study	Cycles analysed up to cycle number	Total number of cycles analysed
Byrd <i>et al.</i>	1987	retrospective	COH-IUI / IUI	all	113	7	7	294
Lalich <i>et al.</i>	1988	retrospective	COH-IUI / IUI	all	138	23	9	553
Dodson <i>et al.</i>	1991	retrospective	COH-IUI / IUI	all	371	7	7	808
Plosker <i>et al.</i>	1994	retrospective	COH-IUI / IUI	all	147	6	5	347
Campana <i>et al.</i>	1996	retrospective	COH-IUI / IUI	male	332	12	9	1086
Berg <i>et al.</i>	1997	retrospective	COH-IUI	male	902	16	9	2908
Hendin <i>et al.</i>	2000	retrospective	not mentioned	all	533	13	9	1673
Khalil <i>et al.</i>	2001	retrospective	COH-IUI	all	893	6	6	2433
Dickey <i>et al.</i>	2002	prospective	COH-IUI	all	1624	7	6	3324
Lee <i>et al.</i>	2002	retrospective	COH-IUI	all	963	4	3	1838
Wainer <i>et al.</i>	2004	retrospective	COH-IUI	all	619	5	5	1549
Iberico <i>et al.</i>	2004	retrospective	COH-IUI	all	402	4	3	924
Dickey <i>et al.</i>	2005	retrospective	COH-IUI	all	2272	7	6	4026
van Rumste <i>et al.</i>	2006	retrospective	COH-IUI	all	300	4	4	879
Dankert <i>et al.</i>	2007	prospective	COH-IUI	male/eci	132	4	4	406
Custers <i>et al.</i>	2008	retrospective	COH-IUI / IUI	all	3714	9	9	15303
Vitthala <i>et al.</i>	2008	retrospective	COH-IUI / IUI	all	397	6	6	1112
de Brucker <i>et al.</i>	2009	retrospective	COH-IUI / IUI	male	1654	24	9	6330
Fahri and Orvieto	2010	retrospective	COH-IUI	all	1035	6	6	2717
Merviel <i>et al.</i>	2010	retrospective	COH-IUI	all	353	9	9	1038
Dorjpurev <i>et al.</i>	2011	retrospective	COH-IUI	all	283	10	9	1055
Groeneveld <i>et al.</i>	2011	retrospective	COH-IUI / IUI	all	378	6	6	1400
Guzick <i>et al.</i>	1986	retrospective	IVF	all	575	9	6	1042
Sharma <i>et al.</i>	1988	retrospective	IVF	all	1295	6	5	2211
Haan <i>et al.</i>	1991	retrospective	IVF	all	1158	4	3	2551
Tan <i>et al.</i>	1992	retrospective	IVF	all	2735	8	6	4736
FIVNAT	1993	retrospective	IVF	all	25666	9	6	58397
Check <i>et al.</i>	1994	prospective	IVF	all	498	3	3	784
Roest <i>et al.</i>	1998	prospective	IVF	all	1211	3	3	2115
Olivius <i>et al.</i>	2002	retrospective	IVF / ICSI	all	974	4	4	1979
Sharma <i>et al.</i>	2002	retrospective	IVF	all	2056	4	4	2709
Witsenburg <i>et al.</i>	2005	prospective	IVF / ICSI	all	750	7	6	1805
Elizur <i>et al.</i>	2006	retrospective	IVF / ICSI	all	1928	14	6	4900
Pelinck <i>et al.</i>	2006	prospective	IVF	all	336	9	3	844
Verhagen <i>et al.</i>	2008	retrospective	IVF / ICSI	all	588	3	3	1024
Malizia <i>et al.</i>	2009	retrospective	IVF / ICSI	all	6164	10	6	14248

COH = controlled ovarian hyperstimulation, IUI = intrauterine insemination, IVF = in vitro fertilization, ICSI = intracytoplasmic sperm injection

TABLE 2. Dropout Rate, Pregnancy Rate and 95 % Confidence Interval (=95% CI) in consecutive IUI cycles

cycle number	Dropout Rate		Pregnancy Rate	
	mean (%)	95% CI	mean (%)	95% CI
1	-	-	12	11,14
2	20	15,24	10	9,12
3	23	18,28	9	7,11
4	31	25,36	8	6,10
5	36	28,43	8	6,11
6	32	24,40	6	4,9
7	48	35,61	9	1,18
8	36	27,44	4	2,7
9	30	22,37	6	1,10

Sharma *et al.*, 2002; Witsenburg *et al.*, 2005; Elizur *et al.*, 2006; Pelinck *et al.*, 2006; Verhagen *et al.*, 2008; Malizia *et al.*, 2009) (Figure 1).

For IUI, 20 studies were retrospective and two were prospective. For IVF 10 studies were retrospective, and four were prospective. The number of treatment cycles per patient reported in a study varied from four to 24 cycles of IUI and three to 14 cycles of IVF. Overall, we analysed 17,555 couples who underwent 52,003 cycles of IUI and 45,934 couples who underwent 99,345 cycles of IVF. The study characteristics are shown in table 1.

For IUI, pregnancy rates ranged from 0% to 50% per cycle (mean 8%; 95% CI 7, 9) and

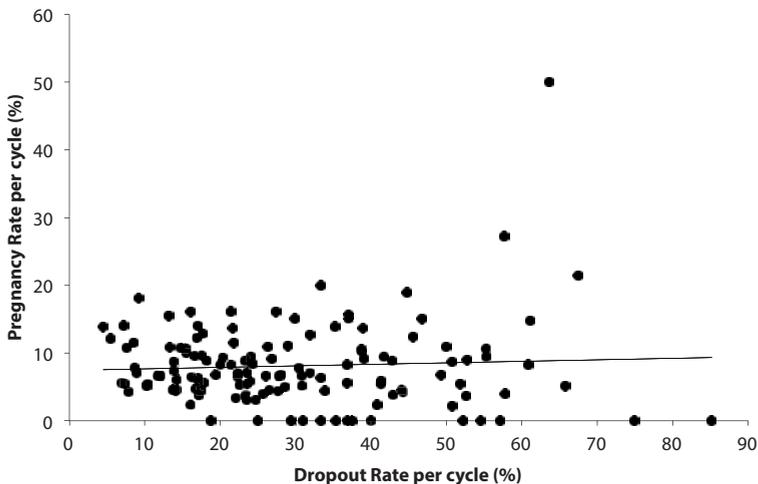


FIGURE 2A. Scatter plot for dropout rates versus pregnancy rates in consecutive IUI cycles. Each dot represents a dropout rate versus a pregnancy rate in the subsequent cycle. The regression line is shown; $y = 0,0223x + 7,4443$. Pearson correlation coefficient 0,058 ($p = 0,50$).

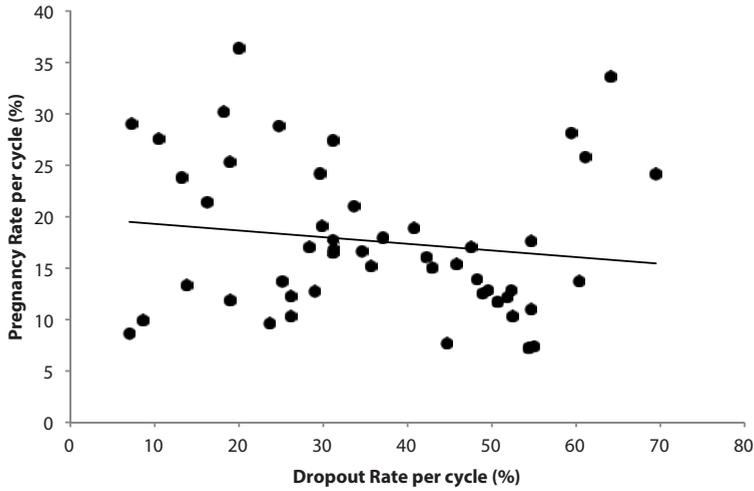


FIGURE 2B. Scatter plot for dropout rates versus pregnancy rates in consecutive IVF cycles. Each dot represents a dropout rate versus a pregnancy rate in the subsequent cycle. The regression line is shown; $y = -0,0647x + 0,1995$. Pearson correlation coefficient $-0,147$ ($p=0,31$).

dropout rates ranged from 4% to 85% (mean 30; 95% CI 27, 33) per cycle. In table 2 shows the mean dropout and pregnancy rates and 95% confidence interval per cycle. The Pearson correlation coefficient for pregnancy rates per cycle in consecutive IUI treatment cycles was $-0,200$ ($p=0,02$). Figure 2A shows a scatter plot of pregnancy rates per cycle as a function of dropout rates per cycle in the previous cycle for consecutive IUI cycles. The Pearson correlation coefficient was $0,058$ ($p=0,50$). Cycle 2 was analysed separately; the Pearson correlation coefficient was $0,240$ ($p=0,26$) as shown in figure 3A.

For IVF, we found pregnancy rates ranging from 7% to 36% (mean 18%; 95% CI 16, 19) per cycle and dropout rates ranging from 7% to 69% per cycle (mean 36%; 95% CI 32, 41). Table 3 shows the mean pregnancy and dropout rates and 95% confidence intervals

TABLE 3. Dropout Rate and Pregnancy Rate with 95% Confidence Interval (=95% CI) per cycle for consecutive IVF cycles

cycle number	Dropout Rate		Pregnancy Rate	
	mean (%)	95% CI	mean (%)	95% CI
1	-	-	22	17,26
2	30	21,40	20	15,24
3	35	26,43	17	14,20
4	45	35,55	18	14,23
5	40	32,48	14	9,18
6	37	29,46	17	15,19

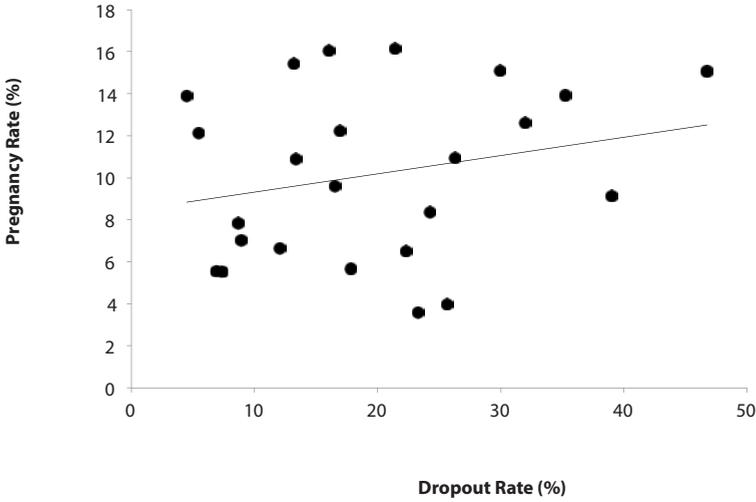


FIGURE 3A. Scatterplot for pregnancy rate in the 2nd cycle and dropout rate in the first cycle of IUI. The regressionline is shown; $y=0,087x+8,4469$. Pearson correlation coefficient 0,240 ($p=0,26$).

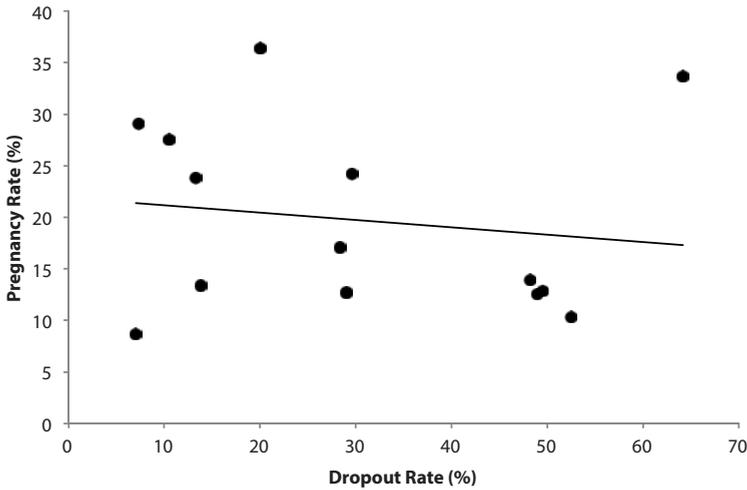


FIGURE 3B. Scatterplot for pregnancy rate in the 2nd cycle and dropout rate in the first cycle of IVF. The regressionline is shown; $y=-0,0713x+21,88$. Pearson correlation coefficient -0,149 ($p=0,61$).

per cycle. The Pearson correlation coefficient for pregnancy rates per cycle in consecutive IVF cycles was -0.297 ($p=0.0002$). Again pregnancy rates were plotted as a function of dropout rates in the previous cycle (figure 2B). The Pearson correlation coefficient was -0.147 ($p=0.31$). The Pearson correlation coefficient for pregnancy rate in cycle 2 versus dropout rate in the previous cycle of treatment was -0.149 ($p=0.61$) as shown in figure 3B.

Discussion

Our results of pooled data from studies reporting on pregnancy rates and dropout rates in consecutive IUI or IVF cycles show no significant impact of dropout rate in a particular cycle on pregnancy rate in the next cycle: a higher dropout rate in one cycle does not result in a higher pregnancy rate of couples continuing treatment in the next cycle. During IUI and IVF treatment, pregnancy rates per cycle decrease significantly in consecutive treatment cycles.

A systematic review of reasons and predictors of discontinuation in fertility treatment found overall that the most frequently occurring reasons for discontinuation were: postponement of treatment (39%), physical and psychological burden (19%), logistics/practical reasons (29%), relational (9%) and personal (17%) problems, treatment rejection (13%) and organizational (12%) or clinic (8%) problems (Gameiro *et al.*, 2012). The reasons for treatment discontinuation were summarized per treatment stage, during the diagnostic fertility workup and before initiation of treatment the most important reasons for discontinuation were relational problems, rejection of treatment in general and financial issues. None of these factors is likely to be predictive for the outcome of IUI, IVF or ICSI. During first-order treatments, like insemination or ovulation induction, reasons for discontinuing treatment were logistical or practical reasons and deferral of treatment. Chances on a pregnancy are equal for these dropouts and those continuing treatment. After the first IVF cycle, censoring by the physician was in 38% of couples the reason to discontinue treatment, other important reasons at this stage were financial issues, psychological or physical burden and clinic related reasons. Physician censoring, usually because of expected poor prognosis, is related to a lower chance of a treatment dependent pregnancy, so dropout of couples for these reasons might inflate pregnancy rates in subsequent treatment cycles (Daya, 2005). Another study showed that 77% of couples chose to withdraw from treatment on their own initiative, while only 23% was advised to stop by their fertility specialist (Brandes *et al.*, 2009). Only a relative small proportion of couples discontinue treatment because of poor prognosis. Another reason for the limited influence of dropout rate on pregnancy rate is possibly the fact that an ART population is a rather homogeneous population. The population is filtered by time and diagnostic tests to specific treatment, e.g. IUI or IVF and treatment success rates of these particular populations vary only in a relative small range. Since homogeneity is a factor that weakens the association between dropout rate and pregnancy rate, only limited influence of dropout rate on pregnancy rate in the next cycle may be expected, as found in this study.

A limitation of our study is that the included studies on IUI and IVF have different definitions of pregnancy, ranging from positive urinary test to live birth. Live birth is the ultimate goal of ART: live birth is the only outcome of interest to the couple treated. Most ART studies report on clinical pregnancy instead of live birth, because follow-up is often not long enough to be able to document the result of a clinical pregnancy. The studies we used also had different endpoints and we did not correct for the endpoints. A previous meta-analysis showed that the effectiveness of treatment based on endpoints as either clinical pregnancy or live birth rate are comparable (Clarke *et al.*, 2009). Other differences in the selected studies are criteria for patient recruitment, the actual treatment given and the amount of cycles performed during ART. The selected studies are ranging from 1986 to 2011 and several adjustments in fertility treatment have been made in the past years. As we did not compare treatment strategies and their results, but the dynamics of dropout on pregnancy rates in the next cycle, we felt free to use the studies for this analysis. Moreover, reimbursement for IVF differed in the selected studies, ranging from full reimbursement in some studies to no reimbursement at all in others. However, also in countries as the Netherlands, where IVF is funded by the health care system up to three ART cycles, dropout rates are high (Brandes *et al.*, 2009).

We conclude that overestimation of pregnancy rates in IUI and IVF treatment due to dropout is limited, even at considerable dropout rates. A high dropout rate in one cycle does not result in a higher pregnancy rate in the next cycle; pregnancy rates in consecutive IUI and IVF treatment cycles are significantly decreasing.

References

- Berg U, Brucker C, and Berg FD, 1997. Effect of motile sperm count after swim-up on outcome of intrauterine insemination. *Fertil Steril* 67:747-750.
- Brandes M, van der Steen JO, Bokdam SB, HAMILTON CJ, de Bruin JP, Nelen WL, and Kremer JA, 2009. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod* 24:3127-3135.
- Byrd W, Ackerman GE, Carr BR, Edman CD, Guzick DS, and McConnell JD, 1987. Treatment of refractory infertility by transcervical intrauterine insemination of washed spermatozoa. *Fertil Steril* 48:921-927.
- Campana A, Sakkas D, Stalberg A, Bianchi PG, Comte I, Pache T, and Walker D, 1996. Intrauterine insemination: evaluation of the results according to the woman's age, sperm quality, total sperm count per insemination and life table analysis. *Hum Reprod* 11:732-736.
- Check JH, Lurie D, Callan C, Baker A, and Benfer K, 1994. Comparison of the cumulative probability of pregnancy after in vitro fertilization-embryo transfer by infertility factor and age. *Fertil Steril* 61:257-261.
- Clarke JF, van Rumste MME, Farquhar CM, Johnson NP, Mol BWJ and Herbison P, 2010. Measuring outcome in fertility trials: can we rely on clinical pregnancy rates? *Fertil Steril* 94:1647-1651.
- Custer IM, Steures P, Hompes P, Flierman P, van Kasteren Y, van Dop PA, van der Veen F and Mol BW, 2008. Intrauterine insemination: how many cycles should we perform? *Hum Reprod* 23:885-888.
- Dankert T, Kremer JA, Cohlen BJ, Hamilton CJ, Pasker-de Jong PC, Straatman H and van Dop, 2007. A randomized clinical trial of clomiphene citrate versus low dose recombinant FSH for ovarian hyperstimulation in intrauterine insemination cycles for unexplained and male subfertility. *Hum Reprod* 22:792-797.
- Daya S, 2005. Life table (survival) analysis to generate cumulative pregnancy rates in assisted reproduction: are we overestimating our success rates? *Hum Reprod* 20:1135-1143.
- de Brucker M, Haentjes P, Evenepoel J, Devroey P, Collins J, Tounaye H, 2009. Cumulative delivery rates in different age groups after artificial insemination with donor sperm. *Hum Reprod* 24:1891-1899.
- Dickey RP, Taylor SN, Lu PY, Sartor BM, Rye PH, and Pyrzak R, 2002. Effect of diagnosis, age, sperm quality, and number of preovulatory follicles on the outcome of multiple cycles of clomiphene citrate-intrauterine insemination. *Fertil Steril* 78:1088-1095.
- Dickey RP, Taylor SN, Lu PY, Sartor BM, Rye PH and Pyrzak R, 2005. Risk factors for high-order multiple pregnancy and multiple birth after controlled ovarian hyperstimulation: results of 4,062 intrauterine insemination cycles. *Fertil Steril* 83:671-683.
- Dodson WC and Haney AF, 1991. Controlled ovarian hyperstimulation and intrauterine insemination for treatment of infertility. *Fertil Steril* 55:457-467.
- Dorjpurev U, Kuwahara A, Yano Y, Taniguchi T, Yamamoto Y, Suto A, Tanaka Y, Matsuzaki T, Yasui T and Irahara M, 2011. Effect of semen characteristics on pregnancy rate following intrauterine insemination. *J Med Invest* 58:127-133.

Eisenberg ML, Smith JF, Millstein SG, Nachtigall RD, Adler NE, Pasch LA and Katz PP, 2010. Predictors of not pursuing infertility treatment after an infertility diagnosis: examination of a prospective U.S. cohort. *Fertil Steril* 94:2369-2371.

Elizur SE, Lerner-Geva L, Levron J, Shulman A, Bider D and Dor J, 2006. Cumulative live birth rate following in vitro fertilization: study of 5,310 cycles. *Gynecol Endocrinol* 22:25-30.

Fahri J and Orvieto R, 2010. Cumulative clinical pregnancy rates after COH and IUI in subfertile couples. *Gynecol Endocrinol* 26:500-504.

French National IVF Registry: analysis of 1986 to 1990 data. FIVNAT (French In Vitro National). *Fertil Steril* 1993;59:587-595.

Gameiro S, Boivin J, Peronace L and Verhaak CM, 2012. Why do patients discontinue fertility treatment? A systematic review of reasons and predictors of discontinuation in fertility treatment. *Hum Reprod Update* 18:652-669.

Groeneveld E, Kouijzer IJ, Timmermans AJ, Schats R and Hompes PG, 2011. Effectiveness of highly purified human menopausal gonadotrophin in intrauterine insemination. *Eur J Obstet Gynecol Reprod Biol* 154:182-186.

Guzick DS, Wilkes C, and Jones HW, Jr, 1986. Cumulative pregnancy rates for in vitro fertilization. *Fertil Steril* 46:663-667.

Haan G, Bernardus RE, Hollanders HM, Leerentveld BA, Prak FM, and Naaktgeboren N, 1991. Selective drop-out in successive in-vitro fertilization attempts: the pendulum danger. *Hum Reprod* 6:939-943.

Hendin BN, Falcone T, Hallak J, Nelson DR, Vemullapalli S, Goldberg J, Thomas AJ, Jr., and Agarwal A, 2000. The effect of patient and semen characteristics on live birth rates following intrauterine insemination: a retrospective study. *J Assist Reprod Genet* 17:245-252.

Ibérico G, Vioque J, Ariza N, Lozano JM, Roca M, Llácer J, Bernabeu R, 2004. Analysis of factors influencing pregnancy rates in homologous intrauterine insemination. *Fertil Steril* 81:1308-1313.

Khalil MR, Rasmussen PE, Erb K, Laursen SB, Rex S, and Westergaard LG, 2001. Homologous intrauterine insemination. An evaluation of prognostic factors based on a review of 2473 cycles. *Acta Obstet Gynecol Scand* 80:74-81.

Lalich RA, Marut EL, Prins GS, and Scommegna A, 1988. Life table analysis of intrauterine insemination pregnancy rates. *Am J Obstet Gynecol* 158:980-984.

Land JA, Courtar DA, and Evers JL, 1997. Patient dropout in an assisted reproductive technology program: implications for pregnancy rates. *Fertil Steril* 68:278-281.

Lee VM, Wong JS, Loh SK, and Leong NK, 2002. Sperm motility in the semen analysis affects the outcome of superovulation intrauterine insemination in the treatment of infertile Asian couples with male factor infertility. *BJOG* 109:115-120.

Malizia BA, Hacker MR and Penzias AS, 2009. Cumulative live-birth rates after in vitro fertilization. *N Engl J Med* 360:236-243.

Merviel P, Heraud MH, Grenier N, Lourdel E, Sanguinet P and Copin H, 2010. Predictive factors for pregnancy after intrauterine insemination (IUI): an analysis of 1038 cycles and a review of the literature. *Fertil Steril* 93:79-88.

Olivius C, Friden B, Borg G, and Bergh C, 2004. Why do couples discontinue in vitro fertilization treatment? A cohort study. *Fertil Steril* 81:258-261.

Olivius K, Friden B, Lundin K, and Bergh C, 2002. Cumulative probability of live birth after three in vitro fertilization/intracytoplasmic sperm injection cycles. *Fertil Steril* 77:505-510.

Pelinc M, Vogel NE, Hoek A, Simons AH, Arts EG, Mochtar MH, Beemsterboer S, Hondelink MN and Heineman MJ, 2006. Cumulative pregnancy rates after three cycles of minimal stimulation IVF and results according to subfertility diagnosis: a multicentre cohort study. *Hum Reprod*. 21:2375-2383.

Plosker SM, Jacobson W, and Amato P, 1994. Predicting and optimizing success in an intra-uterine insemination programme. *Hum Reprod* 9:2014-2021.

Roest J, van Heusden AM, Zeilmaker GH, and Verhoeff A, 1998. Cumulative pregnancy rates and selective drop-out of patients in in-vitro fertilization treatment. *Hum Reprod* 13:339-341.

Sharma V, Allgar V, and Rajkhowa M, 2002. Factors influencing the cumulative conception rate and discontinuation of in vitro fertilization treatment for infertility. *Fertil Steril* 78:40-46.

Sharma V, Riddle A, Mason BA, Pampiglione J, and Campbell S, 1988. An analysis of factors influencing the establishment of a clinical pregnancy in an ultrasound-based ambulatory in vitro fertilization program. *Fertil Steril* 49:468-478.

Smeenk JM, Verhaak CM, Stolwijk AM, Kremer JA, and Braat DD, 2004. Reasons for dropout in an in vitro fertilization/intracytoplasmic sperm injection program. *Fertil Steril* 81:262-268.

Steures P, van der Steeg JW, Mol BW, Eijkemans MJ, van der Veen F, Habbema JD, Hompes PG, Bossuyt PM, Verhoeve HR, van Kasteren YM, van Dop PA, 2004; CECERM. Prediction of an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 82:45-51.

Tan SL, Royston P, Campbell S, Jacobs HS, Betts J, Mason B, and Edwards RG, 1992. Cumulative conception and livebirth rates after in-vitro fertilisation. *Lancet* 339:1390-1394.

van Rumste MM, den Hartog JE, Dumoulin JC, Evers JL and Land JA, 2006. Is controlled ovarian stimulation in intrauterine stimulation an acceptable therapy in couples with unexplained non-conception in the perspective of multiple pregnancies? *Hum Reprod* 21:701-704.

Verhagen TE, Dumoulin JC, Evers JL, Land JA, 2008. What is the most accurate estimate of pregnancy rates in IVF dropouts? *Hum Reprod* 23:1793-9.

Vitthala S, Gelbaya TA, Hunter H, Roberts SA and Nardo LG, 2008. Stimulated intrauterine insemination (SIUI) and donor insemination (DI) as first line management for a selected subfertile population: the Manchester experience. *J Assist Reprod Genet* 25:431-436.

Wainer R, Albert M, Dorion A, Bailly M, Bergere M, Lombroso R, Gombault M, and Selva J, 2004. Influence of the number of motile spermatozoa inseminated and of their morphology on the success of intrauterine insemination. *Hum Reprod* 19:2060-2065.

Witsenburg C, Dieben S, Van der Westerlaken L, Verburg H and Naaktgeboren N, 2005. Cumulative live birth rates in cohorts of patients treated with in vitro fertilization or intracytoplasmic sperm injection. *Fertil Steril* 84:99-107.

PART
III

Ovarian
Reserve

Chapter 6

Value of ovarian reserve testing before IVF: A Clinical decision analysis

Ben Willem J. Mol
Tamara E.M. Verhagen
Dave J. Hendriks
John A. Collins
Arri Coomarasamy
Brent C. Opmeer
Frank J. Broekmans

Human Reproduction 2006;21:1816-1823

Abstract

BACKGROUND: To assess the value of testing for ovarian reserve prior to a first cycle IVF incorporating patient and doctor valuation of mismatches between test results and treatment outcome.

METHODS: A decision model was developed for couples that were considering participation in an IVF program. Three strategies were evaluated: (I) withholding IVF without prior testing, (II) testing for ovarian reserve, and then deciding on IVF treatment if ovarian reserve was estimated to be sufficient, and (III) treatment with IVF without prior ovarian reserve testing.

The outcome considered was the birth of a child. The valuation of the combination of the strategy conducted and the outcome accomplished was expressed on a distress scale in units of 'IVF cycles that were performed in vain'. Correct treatment with IVF and correct withholding of IVF were considered to bring no distress. The distress of withholding IVF in case pregnancy occurred is consequently specified by the ratio of the expected distress after incorrect withholding IVF to the expected distress after incorrect performing IVF (distress ratio). We interviewed both patients and doctors to determine realistic estimates for this distress ratio.

RESULTS: The value of testing for ovarian reserve depends strongly on the expected pregnancy rate after IVF as well as on the valuation of the incorrect decisions from testing. For realistic ranges of the success rate after IVF and for distress ranges as were measured, treatment of all couples without testing was found to generate less distress than testing for ovarian reserve. The sensitivity and specificity of testing for ovarian reserve has to improve to 50% and 96% respectively, to make testing a valuable strategy.

CONCLUSIONS: Based on the decision analysis, where current test accuracy and preference inventory among patients and physicians were used, testing for ovarian reserve seems not useful for current IVF programmes.

Introduction

The ability of the ovaries to respond to exogenous gonadotropins stimulation and to develop several follicles with subsequent embryo selection after IVF of retrieved oocytes is one of the success factors in IVF. The occurrence of a poor response, often with cancellation of the treatment cycle, is a common problem. Identification of potential poor responders before the actual start of the IVF treatment would potentially enable physicians to counsel these patients on the prospects of the treatment that they are about to begin and even to actively withhold treatment in certain cases.

We have previously reported on the accuracy of several tests to assess the ovarian reserve prior to IVF treatment. Among these tests, the Antral Follicle Count (AFC) has been shown to be a significantly better predictor of poor response than basal FSH (Hendriks *et al.*, 2005). However, all available tests to date perform inaccurately in the prediction of pregnancy.

The clinical question of whether ovarian reserve testing before IVF has more benefit than harm depends not only on their discriminative accuracy, but also on the consequences coupled to the test result and the valuation of the test-based consequences. Net benefit and harm of a strategy that incorporates ovarian reserve testing before IVF need to be compared to other clinically relevant strategies. An attractive strategy, which is explicitly or implicitly used in current clinical practice, is the use of the actual response in the first stimulation cycle. A poor response could imply discontinuation of treatment, as the chance of IVF-success would be approaching zero. Recent work has shown that only the combination of an abnormal ovarian reserve test and a poor ovarian response in the first cycle (the expected poor response case) indicates very low prospects in subsequent cycles (Klinkert, *et.*, 2004).

Although currently available tests have only very limited predictive capacity for pregnancy, a combination of existing tests in multivariable models or the introduction of new tests could resolve this issue. However, the requirements of such composite or new tests in terms of diagnostic accuracy in order to be useful in clinical practice are unknown. In this paper, we use a decision analytic framework, modelling the test accuracy as well as patients and doctors perception of these outcomes. Value judgments of both incorrect start of IVF (i.e. IVF without a pregnancy) and incorrect withholding of IVF (i.e. not starting IVF where pregnancy would have occurred) are inventoried among a group of patients and doctors working in the field of infertility. In a decision analysis we evaluated whether the expected benefits based on probabilities and value judgments of outcomes would justify testing for ovarian reserve or treatment without prior testing.

Material and methods

General problem definition and analytical approach

We used a representation of the decision problem in terms of three available options, analogous to the approach described by Pauker and Kassirer (1980): (I) withholding IVF treatment, (II) test for ovarian reserve and then decide for IVF if ovarian reserve is judged sufficient and (III) treatment without prior testing for ovarian reserve. The decision whether to treat couples without prior testing, to withhold treatment or to test for ovarian reserve and to treat in case of a predicted normal response is based on the expected value of each alternative. The expected value of an alternative depends on the probability of possible outcomes when applying the strategy and the subjective values attached to the outcomes in relation to the chosen strategy. A decision tree was constructed, in which we modelled these three alternative strategies for a subfertile couple that are about to consider treatment with IVF, and Treeage decision analysis software (DATA) was used for performing the decision analyses. The clinical outcome was ongoing pregnancy resulting in the birth of at least one living baby. We will refer to this outcome as ‘pregnancy’ throughout the manuscript.

Treatment strategies

We compared the three mutually exclusive alternatives, (I) withhold treatment without testing for ovarian reserve, (II) test for ovarian reserve, and then decide for IVF treatment in case ovarian reserve was judged to be sufficient, and (III) treatment of the couple without prior testing for ovarian reserve and withholding further treatment in case of poor response (Figure 1).

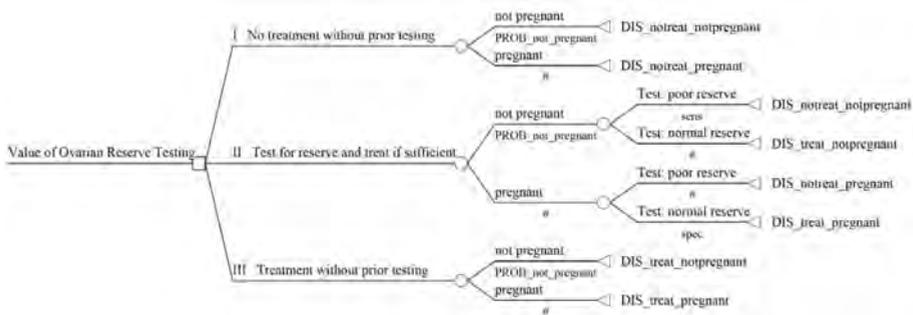


FIGURE 1. Decision tree representing the three strategies, the probabilities of different outcomes and the values (in terms of distress (DIS=distress)) attached to these outcomes.

	Would not have conceived after an IVF-attempt	Would have conceived after an IVF-attempt
Test shows poor ovarian reserve, and IVF is postponed	No distress	Distress
Test shows adequate ovarian reserve, and IVF is performed	Distress	No distress

FIGURE 2. Schematic view of the four possible outcomes of the decision tree, and their valuation. The patients with a test indicating normal ovarian reserve that would have conceived after IVF, and those with a test indication poor ovarian reserve and would not have conceived, even if IVF had been performed, have no distress and their valuation of the outcome is considered to be zero. The other two outcomes, i.e. no pregnancy after an IVF-attempt in a patient in whom the test indicated normal response, or no pregnancy after withholding IVF since the test indicated poor ovarian response, but in which IVF would have been successful when it would have been performed, are valued on a negative scale. The value of no pregnancy after an IVF attempt is considered to be the standard, whereas the value of incorrect withholding IVF is determined by the distress ratio, i.e. how much worse is withholding IVF as compared to performing IVF in vain.

Outcomes

The foremost relevant clinical outcome to evaluate is the extent to which different strategies result in ongoing pregnancies. However, we hypothesize that from the perspective of the couple, a pregnancy that fails to occur after IVF is valued differently from a couple not becoming pregnant without IVF. Four possible outcomes are (a) correct start of IVF (i.e. start of an IVF cycle resulting in pregnancy), (b) incorrect start of IVF (i.e. start of an IVF cycle not resulting in pregnancy), (c) incorrect refraining from IVF (i.e. withholding an IVF cycle that would have resulted in pregnancy) and (d) correct refraining from IVF (i.e. withholding an IVF cycle that would not have resulted in pregnancy) (Figure 2). These outcomes will be further referred to as ‘Correct IVF’, ‘Incorrect IVF’, ‘Incorrect No IVF’ and ‘Correct No IVF’ respectively. In order to appropriately evaluate different strategies, the subjective valuation of these outcomes of the test and treatment cycle and the probabilities that they occur have to be taken into account.

The subjective value that can be attached to each of the four possible outcomes can be expressed in terms of the expected distress that is produced from incorrect decisions on starting or withholding IVF treatment. Distress is the factor that represents the distress that emerges from the actual knowledge that the policy has been wrong and can be expressed on a distress scale. The correct start of treatment (‘Correct IVF’) and the correct withholding of treatment (‘Correct No IVF’) both bring no additional distress, whereas the other two possible combinations, incorrect starting or withholding of treatment indeed will generate distress. The central issue here concerns the distress of ‘Incorrect

no IVF' relative to 'Incorrect IVF'. In other words, how much more is the couple assumed to suffer from incorrect withholding IVF as compared to incorrect starting of IVF. This ratio of the distress of 'Incorrect No IVF' and the distress of 'Incorrect IVF' is referred to as the 'distress ratio' (Thornton and Lilford 1995; Van der Meulen *et al.*, 1999). The unit in which distress for 'Incorrect No IVF' is expressed and can be interpreted is 'incorrect start of IVF equivalent'. One 'incorrect start of IVF equivalent' corresponds with the distress of the incorrect start of IVF in one treatment cycle.

The expected distress for each strategy is calculated from the probabilities of each outcome weighed by the value (distress) attached to that outcome (calculation formulas are provided in the Appendix). According to the principles of decision theory, the best choice is the one with the lowest expected distress. The decision whether to withhold treatment, to test for ovarian reserve and treat with IVF if sufficient or to treat with IVF without testing, thus depends on the expected distress of each strategy.

Inventory of preferences among patients and doctors

To make an inventory of the preferences, we performed face-to-face interviews among subfertile couples that were scheduled for IVF in the Utrecht Medical Centre in Utrecht, The Netherlands. If indicated, IVF is offered without restrictions to couples in which the female's age is < 41 years. In the interview, couples received a written explanation on the use of a fictive prediction test prior to IVF designed to decide whether couples would be allowed or not to enter the treatment. Also, the imperfections of testing were discussed and written information was further explained by one of the authors (DH). Subsequently, they were asked how much worse they would value the incorrect withholding of IVF based on this fictive test as compared to the incorrect start of IVF based on the same test. For the inventory of preferences among physicians working in the field of reproductive medicine, we approached gynaecologists and fertility doctors in six other clinics performing IVF in the Netherlands, and presented them a situation in which a fictive test was done prior to IVF that would allow or not a couple starting IVF treatment. Again, they were asked how much worse they would value the incorrect withholding of IVF based on this fictive test as compared to the incorrect start of IVF based on the same test.

Decision analyses

The following parameters, based on a review of the literature, are used in the decision analyses (Table 1). The probability of non-pregnancy in a first cycle IVF was set at 80%, sensitivity and specificity at 24% and 90% respectively, and the distress ratio was set at 100, i.e. the incorrect withholding of IVF was judged 100 times worse than the incorrect

TABLE 1. Baseline Parameter Values, Parameter Ranges Used in Sensitivity Analysis, and References Used

Parameter	Variable	Baseline	Range	Reference
probability of non-pregnancy	PROB_pregnant	0.80	0.40-1.00	
Sensitivity	sens	0.24		Hendriks et al., 2005
Specificity	spec	0.90		Hendriks et al., 2005
distress ratio	DIS_no treat_pregnant*	100	10-250	

*DIS_treat_notpregnant is set at 1

starting of IVF. The expected distress for each of the strategies was then calculated by weighing the disutilities associated with each outcome with the probability of that outcome (see Appendix).

Subsequently, we calculated the expected distress for a wide range of prevalence of non-pregnancy (40% to 100%) and a wide range of distress ratios (10 to 250). A distress ratio of 5 was excluded because we gave more importance to the patients' distress than the doctors' distress on their behalf, and 1000 was excluded because it was ridiculously high. We then plotted the expected distress as a function of the probability of non-pregnancy. From these graphs, we were able to derive the strategy that would result in the lowest expected distress.

The accuracy required to make testing a better strategy than treatment of all couples is evaluated by assessing the threshold for the prevalence of non-pregnancy at which testing for ovarian reserve becomes superior to treatment of all couples. In other words, we should focus on the threshold for the prevalence of non-pregnancy at which treating without testing results in an expected distress that is equal to that of testing with subsequent treatment of those with expected normal ovarian reserve (Pauker and Kassirer, 1980). If the prevalence of non-pregnancy is lower than this threshold, treatment of all couples should be the preferred treatment, whereas otherwise a strategy starting with testing for ovarian reserve would be preferable.

We calculated the expected distress for each of the three strategies for each combination of sensitivity and specificity, when prevalence of non-pregnancy and distress-ratio are fixed. The strategies I, treat nobody without testing, and III, treat all without testing, result in a fixed expected distress that is independent of sensitivity and specificity. For strategy II, test all couples for ovarian reserve, and then decide on treatment, the expected distress will be minimal in case the test characteristics sensitivity and specificity are perfect (i.e. both 100%). This distress increases once test properties decrease. The

threshold is then determined by that combination of sensitivity and specificity at which the expected distress of strategy II, testing for ovarian reserve, is equal to the lowest of the two other strategies. This threshold has been plotted into a Receiver-Operating-Characteristic (ROC)-space for different distress ratios for a probability of non-pregnancy of 80% and 50%.

Results

Inventory of preferences among patients and doctors

We performed an inventory among patient couples that were indicated to initiate IVF treatment. For these couples a structured interview was designed, including a description of the use and characteristics of a given, fictive, ovarian reserve test, and the question to value the incorrect withholding of IVF as compared to the incorrect start of IVF. In total 25 couples were asked to participate of which 20 (80%) agreed and finished the structured interview. The valuations by couples resulted in distress ratios ranging from 50 to 1000, with a median of 250, indicating that couples in general expected 250 times more distress by incorrect refusal of IVF than starting IVF without getting pregnant.

For the 36 clinicians working in the field of assisted reproduction we prepared a written survey in which they were asked to value incorrect withholding of IVF as compared to an incorrect start of IVF. Surveys were then sent to representatives of all 7 IVF centres in the

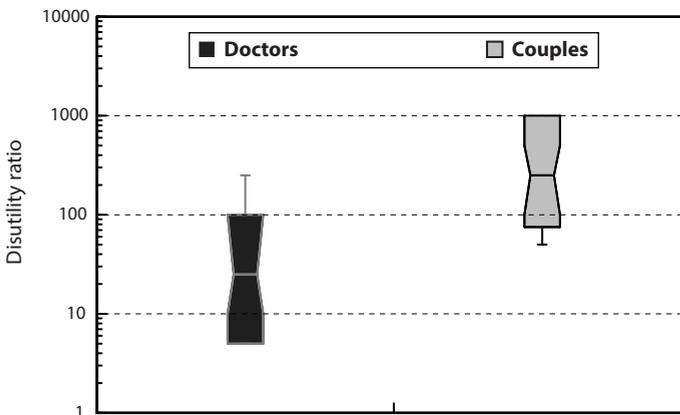


FIGURE 3. Box-and-whisker plot showing the distribution of the distress ratio according to doctors and patients. The distress ratio represents how much worse incorrect withholding IVF is values as compared to incorrect starting of IVF. For example, a distress ratio of 100 implicates that not starting IVF in case a pregnancy would have occurred is values 100 times as worse as the start of an IVF cycle that did not result in pregnancy.

Netherlands (n=36). The response rate was 70%. From the 25 responding clinicians, one felt unable to answer the questions posed. From the answers of the other 24 the distress ratios ranged between 5 and 250, with a median of 25 (Figure 3). The difference between the distress ratios recorded among physicians and patient couples, was statistically significant (P-value < .01, Mann-Whitney U-test). The results obtained in this inventory were used as baseline and range parameters in the subsequent decision analyses.

Decision analyses

In the baseline analysis, the expected distress for the three strategies are 20, 2.6 and 0.80 'incorrect start of IVF equivalents' for strategies I, II and III respectively, indicating that treatment of couples without prior testing (strategy III) generates the least distress. In subsequent analyses, the probability of non-pregnancy was varied between 40% and 100%, and sensitivity and specificity remained constant at their baseline values (24% and 90%, respectively), corresponding with the results of a previous meta-analysis on the accuracy of AFC (Hendriks *et al.* 2005). Figure 4A shows the expected distress for the three strategies, when the distress ratio is presumed to be 100, i.e. the incorrect withholding of IVF is valued 100 times as distressing as the incorrect performance of IVF. When the probability of non-pregnancy was below 77% (i.e. an IVF success rate of 23% per cycle or higher), strategy III had the lowest expected distress. In case of a probability of non-pregnancy between 97.5% and 99.5% (i.e. IVF success rates of 2.5% to .5% per cycle), the strategy with testing for ovarian reserve (strategy II) was expected to have the lowest expected distress, whereas above a probability of non-pregnancy of 99.5% (IVF success rate per cycle lower than .5%), no treatment (strategy I) would bring the lowest expected distress.

Figure 4B, C and D depict the situation for distress ratios of 10, 50 and 250, respectively, showing that with the incorrect withholding of IVF distress increases much more rapidly with increasing distress ratio compared to incorrect application of IVF treatment. These figures show the same pattern as for a distress ratio of 10, although the 'treat all-to-test all' threshold and the 'test all-to-no treatment' threshold both shift to higher probabilities of failure of IVF. The probability of IVF-failure at which testing brings less distress than the other two strategies lies between 77% and 93% for a distress ratio of 10, 94% and 98.5% for a distress ratio of 50, and between 99% and 99.7% for a distress ratio of 250.

Subsequently we addressed the question to which degree test accuracy should be improved to make testing a better strategy than treatment of all couples without testing by determining those levels of sensitivity and specificity at which testing for ovarian reserve will become superior to treatment of all couples without testing. Instead of

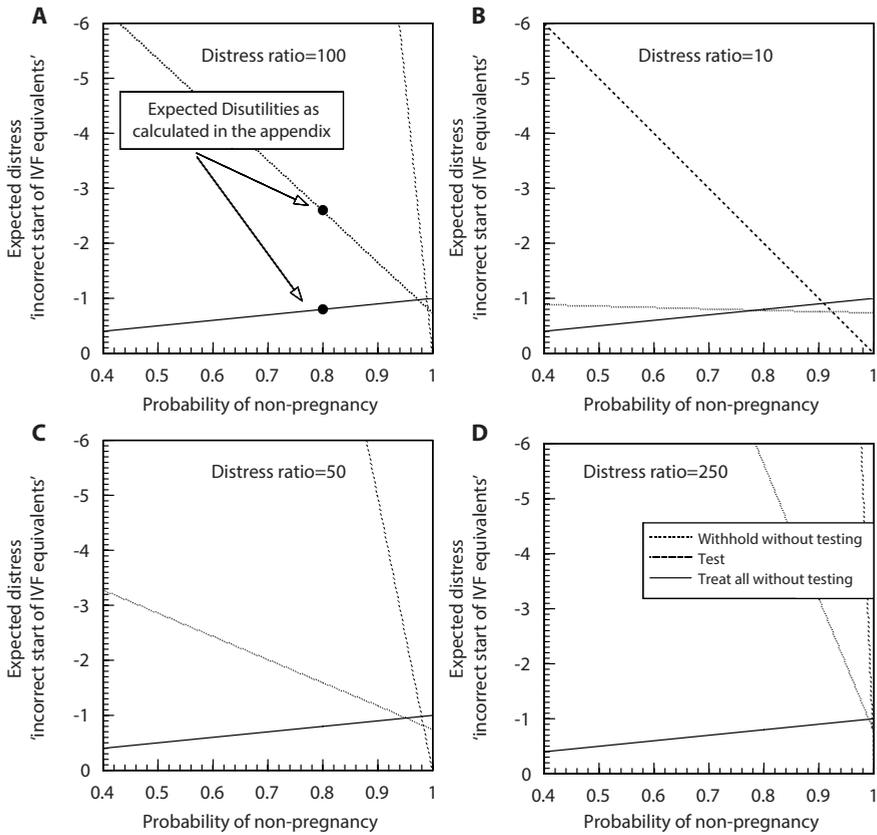


FIGURE 4AD. Expected distress for three strategies

- (I) withhold treatment without testing for ovarian reserve
- (II) test for ovarian reserve, and then decide for IVF treatment in case ovarian reserve was judged to be sufficient
- (III) treat the couple without prior testing for ovarian reserve

In each figure, the prevalence of non-pregnancy is varies between 40% and 100%, and the expected distress is calculated for distress ratios between 10 and 250. The distress ration represents the distress caused by incorrect withholding of IVF as compared to the distress caused by incorrect starting IVF. For example, a distress ratio of 10 implicates that not starting IVF in a situation where pregnancy would have occurred would be valued 10 times as worse as starting IVF in a situation where pregnancy would not have occurred.

keeping sensitivity and specificity constant at their baseline values, prevalence of non-pregnancy after IVF was fixed at 80% and the distress ratio was set at 10 (i.e. not starting IVF in a situation where pregnancy would have occurred would be valued 10 times as distressing as starting IVF in a situation where pregnancy would not have occurred). Figure 5A shows a ROC curve for this situation, indicating that testing would be superior over treatment of all couples when the accuracy of ovarian reserve testing is better than the line between the left lower corner (0% sensitivity and 100% specificity) and the

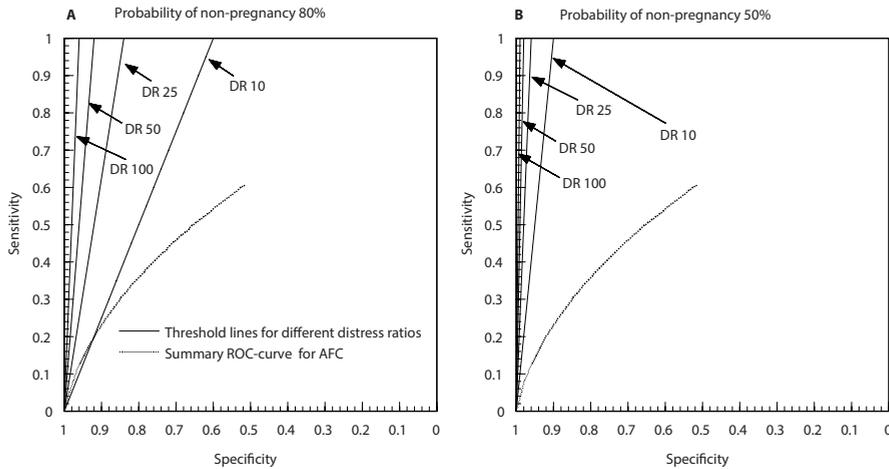


FIGURE 5AB. Receiver-Operating-Characteristic space for prevalences of non-pregnancy of 80% (figure 5A) and 50%, respectively. The straight lines represent the threshold at which an ovarian reserve test becomes accurate enough to become more valuable than treatment of all couples without testing. Combinations of sensitivity and specificity left of these lines represent test combinations that are good enough to justify testing. As can be seen from the figures, a higher distress ratio and a lower prevalence of non-pregnancy make immediate treatment more attractive than testing. The figures also show a summary ROC-curve for AFC in the prediction of non-pregnancy, as was estimated in a meta-analysis (Hendriks *et al.*, 2005). This ROC curve is only good enough if the prevalence of non-pregnancy is high, and if the distress ratio is below 10. DR = distress ratio

combination of 100% sensitivity and 60% specificity. The point at which sensitivity is 0% and specificity is 100% implicates that at this point none of the cases of diminished ovarian reserve are detected, and that all couples are therefore treated. This test strategy has therefore at this combination of sensitivity and specificity an expected distress that is equal to the immediate treatment of all couples (-0.8 'incorrect start of IVF equivalent'). Once specificity subsequently decreases, the sensitivity has to increase in order to keep the expected distress equal to '-0.8 'incorrect start of IVF equivalent'', which is the expected distress of treatment of all couples. For a sensitivity of 100%, this is the case at a specificity of 60%.

As can also be seen from this ROC diagram, only a small part of the summary ROC curve for the AFC lies to the left of this threshold line. This implies that only with the use of an extreme cut-off for an abnormal test the desired level of expected distress is realized. Figure 5A also shows that with the increase of the distress ratio to 25, 50 and 100, (i.e. the valuation of incorrect withholding of IVF as more distressing), the demands placed upon the test accuracy sharply rise. As a result, for distress ratios of 25 and higher, testing for ovarian reserve is always inferior to treatment of all couples without testing.

Figure 5B shows the situation for a probability of non-pregnancy of 50%. As can be seen

from this ROC curve, the demands put upon test accuracy increase in this situation to such an extent that treatment of all couples is always superior to testing, even if the distress ratio is as low as 10.

Discussion

This study shows that the value of ovarian reserve testing prior to IVF for individual couples strongly depends on the prevalence of IVF failure as well as on the valuation of the false positive (incorrect withholding IVF) or false negative (incorrect performing IVF) outcomes. According to the decision model used in this study, testing for ovarian reserve is only worthwhile when the distress ratio does not exceed 10, and the prevalence of non-pregnancy is around 80%. When we performed an inventory of these preferences among physicians in the field of reproductive medicine and IVF-indicated patient couples, patients expected to suffer far less distress from IVF that was incorrectly started than physicians thought they would. Although we did not evaluate test accuracy but rather the valuation of false-positive and false-negative outcomes, these data show that the clinical value of available tests, given their present accuracy level, is low and that the routine use of these tests in clinical practice can be reasonably questioned.

Whether individual differences in rating the true outcome as compared to the outcome predicted by a test can affect the value of that test is only rarely addressed. In the vast majority of test evaluations, predetermined threshold risks are used, above which testing is advised as clinically useful and offered to the patient. As explained in the Methods section, the threshold risk between starting IVF and testing for ovarian reserve on one hand and between testing for ovarian reserve and not starting IVF on the other hand is fully determined by the distress ratio (Van der Meulen *et al.* 1999). At a prevalence of non-pregnancy of 77%, which is in our opinion a rather modest estimate of the success rate of IVF per cycle in present programs, the distress ratio of 10 resulted in equal expected distress between treatment of all couples and testing for ovarian reserve. The interviews indicated that median preferences of both patients and doctors were far above this distress ratio. This implies that from both the patient and physician point of view not any ovarian reserve test should be used for the decision whether or not to start IVF. Instead, for counselling purpose or for adjustment of treatment schedules, the test may elicit much less disappreciation of false-positive test results and hence be more clinically acceptable

and applicable. However, if testing only implies such minor consequences then health improvement or efficiency gain may become too low to justify the burden of testing.

Decision analysis for diagnostic issues is more complex than for therapeutic issues. In therapeutic issues, the outcome measure is usually unidimensional, as the patient does or does not benefit from treatment. In diagnostic issues, the outcome is influenced both by the discriminatory performance of the test under study and by the effects of subsequent treatment as compared to no treatment. Thus, there are four possible outcomes, i.e. true-positive test results, false-positive test results, true-negative test results and false-negative test results. We decided to assign no distress to consequences of true-positive results and true-negative results, since these two outcomes represent the optimal result from a diagnostic point of view. The other two consequences of test results were then valued on a negative scale, and combined in a distress ratio.

Variations in the pregnancy rate of the IVF program have an important effect on the value of the ovarian reserve tests. If the pregnancy rate increases from 20% to 50%, the test accuracy of ovarian reserve testing has to improve very strongly, as can be seen from the comparison between figure 5A and figure 5B. As the comparison with the summary ROC curve for the AFC shows, its test accuracy is too low to be of clinical value, even at distress ratios as low as 10.

In the field of reproductive medicine, recently a shift was made from surrogate outcomes such as ovarian response, ovum pick-up rate and embryo-transfer rate towards more relevant outcomes, of which live birth rate is the most important one. In the literature on testing for ovarian reserve, most studies report on ovarian response rate and/or pregnancy rate. Since pregnancy rate is the outcome most relevant for the patient, we used that one in our study.

The number of performed cycles was not specified in the decision analysis. One could speculate that if the observed outcome (pregnancy) would be calculated over a series of IVF cycles, the distress ratio would become decreased, since three cycles of IVF performed without a pregnancy will be valued worse than one cycle of IVF without a pregnancy. On the other hand test performance may well improve if the predicted outcome may become realized in more than one attempt, as more exposures give better opportunity to separate the reproductive successful cases from those that are bound to fail. Three cycles of IVF will correspond with an expected pregnancy rate of 50%. As Figure 5B shows, the distress

ratios have to be low to make ovarian testing potentially of value. In view of the high distress ratio that we measured, it is unlikely that patients will change these valuations strongly when three cycles are performed instead of one. Only if the performance curve for the test would clearly shift to the left and the valuations would change to such an extent that the threshold lines would fan to the right testing may become useful.

The technique of using a ratio to assign values to outcomes of a medical procedure is known as magnitude estimation (Stevens 1971; Froberg and Kane 1989). An argument for its use is that it is very easy to ask patients to quantify how much worse they think that the not performing IVF would be compared with performing IVF in vain. Such a ratio represents the ovarian reserve testing decision as a choice between certain outcomes and therefore does not account for the attitude of patients to risk (Watson and Buede 1987). As an alternative to measuring the values of the outcomes of the IVF programme, which incorporates this risk attitude, one could use the more complex standard reference gambling technique, where values are measured through the elicitation of a series of preferences between lotteries. Most people are averse to risk, which means that the distress ratios used in our study might underestimate the expected distress after withholding IVF incorrectly, especially if the distress ratio is relatively high. Empirical research is required to investigate to what extent the use of standard reference gambles would result in different decisions regarding testing for ovarian reserve.

We conclude that the value of the testing for ovarian reserve should also be determined on the basis of individual preferences, and that the use of a predetermined threshold risk above which IVF is withheld should not be the only factor that directs the decision on whether or not to use a certain test. The distress ratio as introduced in this study may be an instrument to incorporate subjective judgement of the outcomes of an IVF programme for individual couples in clinical practice. As our interviews demonstrated, elicitation of these values is possible in clinical practice, since couples answered the question in a more or less consistent way. Moreover, their preferences were clearly different from that of the doctors, thus indicating that doctors should incorporate the preference of their patients in their decisions, rather than deciding for them. The result of this study supports the concept of formal decision making in individual couples to complement the process of counselling.

References

Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: scaling methods. *J Clin Epidemiol* 1989; 42: 459-71.

Hendriks DJ, Mol BWJ, Bancsi LFJMM, Te Velde ER, , Broekmans FJM. The accuracy of the antral follicle count and basal FSH in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. *Fertil Steril*. 2005; 83: 291-301.

Klinkert ER, Broekmans FJ, Looman CW, te Velde ER A poor response in the first in vitro fertilization cycle is not necessarily related to a poor prognosis in subsequent cycles. *Fertil Steril*. 2004 ;81:1247-53.

Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302:1109-17

Stevens SS. Issues in psychophysical measurement. *Psychol Rev* 1971; 78: 426-50.

Thornton JG, Lilford RJ. Decision analysis for medical managers. *BMJ* 1995; 310: 791-4.

Van der Meulen JH, Mol BW, Pajkrt E, van Lith JM, Voorn W. Use of the disutility ratio in prenatal screening for Down's syndrome. *Br J Obstet Gynaecol*. 1999 ;106:108-15

Watson SR, Buede DM. Decision synthesis. The principles and practise of decision analysis. Cambridge University Press, Cambridge 1987.

Appendix

calculation of the expected distress for each of the strategies.

For strategy (I), treat none of the couples without testing, the expected distress can be calculated from:

$$(1A) \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{notreat_notpregnant}} + (1 - \text{Prob}_{\text{notpregnant}}) * \text{Exp_DIS}_{\text{notreat_pregnant}}$$

As withholding treatment in women who do not become pregnant ($\text{Exp_DIS}_{\text{notreat_notpregnant}}$) is assumed to bring no distress, this can be simplified to:

$$(1B) (1 - \text{Prob}_{\text{notpregnant}}) * \text{Exp_DIS}_{\text{notreat_pregnant}}$$

With the baseline parameter for prevalence of not becoming pregnant set at 80%, and the expected distress of withholding treatment with IVF in case of normal ovarian reserve equal to 100, the expected distress of this strategy will be -20 'incorrect start of IVF' equivalents' (not shown in Figure 4, as the Y-axis is scaled to -6).

For strategy (II), test all couples for ovarian reserve, and then decide on treatment, the expected distress can be calculated from:

$$(2A) \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{notreat_notpregnant}} * \text{Sensitivity ovarian reserve test} + \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{treat_notpregnant}} * (1 - \text{Sensitivity ovarian reserve test}) + (1 - \text{Prob}_{\text{notpregnant}}) * \text{Exp_DIS}_{\text{treat_pregnant}} * (\text{Specificity ovarian reserve test}) + (1 - \text{Prob}_{\text{notpregnant}}) * \text{Exp_DIS}_{\text{notreat_pregnant}} * (1 - \text{Specificity ovarian reserve test}).$$

As the start of treatment in women who will conceive ($\text{Exp_DIS}_{\text{treat_pregnant}}$) and withholding treatment in women who will not conceive ($\text{Exp_DIS}_{\text{notreat_notpregnant}}$) both are expected to bring no distress, this can be simplified to:

$$(2B) \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{treat_notpregnant}} * (1 - \text{Sensitivity ovarian reserve test}) + (1 - \text{Prob}_{\text{notpregnant}}) * \text{Exp_DIS}_{\text{notreat_pregnant}} * (1 - \text{Specificity ovarian reserve test}).$$

If the prevalence of not becoming pregnant is set at 80%, the expected distress of incorrect starting treatment with IVF is set at 1, the expected distress of incorrect withholding treatment with IVF is 10, and the sensitivity and specificity of ovarian reserve testing equals 24% and 90%, respectively, then the expected distress of this strategy will be -0.80 'incorrect start of IVF' equivalents' (shown in Figure 4 as black dot).

For strategy (III), treat all couples without testing and decide on further treatment based on first cycle ovarian response the expected distress was calculated from:

$$(3A) \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{treat_nonpregnancy}} + (1 - \text{Prob}_{\text{not_pregnant}}) * \text{Exp_DIS}_{\text{treat_pregnant}}$$

As the start of treatment in women that will conceive (Exp_Utreat_pregnant) is assumed to bring no distress (i.e. Exp_DIS set a zero), this can be simplified to:

$$(3B) \text{Prob}_{\text{notpregnant}} * \text{Exp_DIS}_{\text{treat_notpregnant}}$$

If the prevalence of not becoming pregnant after IVF-ET is set at 80%, and an expected distress of treatment with IVF in case of not becoming pregnant is 1, then the expected distress of this strategy will be - 2.6 'incorrect start of IVF equivalents' (shown in Figure 4 as black dot).

Chapter 7

The accuracy of multivariate models predicting ovarian reserve and pregnancy after In Vitro Fertilization: a meta-analysis

Tamara E.M. Verhagen

Dave J. Hendriks

Laszlo F.J.M.M. Bancsi

Ben Willem J. Mol

Frank J.M. Broekmans

Human Reproduction Update 2008;14:95-100

Abstract

BACKGROUND: To review the accuracy of multivariate models for the prediction of ovarian response and pregnancy in women undergoing IVF compared with the antral follicle count (AFC) as single test.

METHODS: We performed a computerised MEDLINE and EMBASE search to identify articles published on multivariable models for ovarian reserve testing in patients undergoing IVF. In order to be selected, articles had to contain data on the outcome of IVF in terms of either pregnancy and/or poor response and on the prediction of these events based on a multivariate model. For the selected studies, sensitivity and specificity of the test in the prediction of poor ovarian response and non-pregnancy were calculated. Overall performance was assessed by estimating a summary Receiver Operating Characteristic (ROC) curve, which was compared to the ROC curve for the AFC as the current best single test.

RESULTS: We identified 11 studies reporting on the predictive capacity of multivariate models in ovarian reserve testing. All studies reported on the prediction of poor ovarian response, whereas none reported on the occurrence of pregnancy. The sensitivity for prediction of poor ovarian response varied between 39% and 97% and the specificity between 50% and 96%. Logistic regression analysis indicated that cohort studies provided a significantly better discriminative performance than case-control studies. As cohort studies are superior to case-control studies further analysis was limited to the cohort studies. For the cohort studies, a summary ROC curve could be estimated, which had a similar shape as a previously made summary ROC curve for the AFC.

CONCLUSIONS: The accuracy of multivariate models for the prediction of ovarian response in women undergoing IVF is similar to the accuracy of AFC. No data are available on the capacity of these models to predict pregnancy, let alone live birth. On the basis of these findings, the use of more than one single test for the assessment of ovarian reserve cannot currently be supported.

Introduction

In recent decades, societal changes in family planning have caused a significant increase in the incidence of unwanted infertility due to female reproductive aging (Weinstein *et al.*, 1993; Abma *et al.*, 1997). Ageing of women leads to more female fertility problems because of a reduction in primordial follicle pool and loss of oocyte quality (te Velde and Pearson, 2002). An age-related effect on female fecundity has also been shown in reports on the results of IVF treatment (Templeton *et al.*, 1996; Quban *et al.*, 2005). In an IVF programme, ovarian ageing is characterised by decreased ovarian responsiveness to gonadotrophin administration and lowered pregnancy rates. Identification of women at increased risk for poor ovarian response prior to IVF could be useful, as this either could prevent unnecessary continuation of treatment (Klinkert *et al.*, 2004), or result in adjustment of the gonadotrophin dose in order to maximise ovarian response (Popovic-Todorovic *et al.*, 2003; Klinkert *et al.*, 2005).

In a recent systematic review of tests predicting ovarian reserve and IVF outcome, we concluded that none of the individual ovarian reserve tests (ORT) can routinely be used for outcome prediction (Broekmans *et al.*, 2006). ORT evaluated were early follicular phase blood values of FSH, estradiol, inhibin B and anti-Mullerian Hormone (AMH), antral follicle count (AFC), ovarian volume (OV) and ovarian blood flow. Furthermore, the clomiphene citrate challenge test (CCCT), the exogenous FSH ovarian reserve test and the gonadotrophin agonist stimulation test (GAST) were evaluated. The capacity of each of the evaluated tests to predict ovarian reserve was modest to poor.

When patient and doctor valuation of mismatches between test results and treatment outcome are taken into account, testing for ovarian reserve seems not useful for current IVF programmes (Mol *et al.*, 2006). In view of these results, it is currently recommended to start IVF treatment without any prior testing.

Combining tests that predict the ovarian reserve might lead to a better estimation of the ovarian reserve capacity. If this were to be the case, testing for ovarian reserve prior to IVF might become more useful. Our group has previously reported on the performance of tests for ovarian reserve prior to IVF in a multivariate setting (Bancsi *et al.*, 2002; van Rooij *et al.*, 2002; Hendriks *et al.*, 2005b). These individual studies showed that the combination of tests leads to a better prediction than individual tests. In a previously published meta-analysis, AFC and basal FSH were compared as predictors of ovarian reserve and this meta-analysis showed a significantly better predictive performance

of AFC (Hendriks *et al.*, 2005b). An estimated summary ROC curve was constructed for AFC in that study. A preliminary assessment of the literature on the subject has been performed (Broekmans *et al.*, 2006). In this article we report on a systematic review and meta-analysis on the subject. The aim of this meta-analysis was to assess the accuracy of a multivariate approach of ovarian reserve testing. We compared the multivariate prediction models of ovarian reserve testing with a previously study on univariate prediction, were we concluded that AFC might be considered the test of first choice in the assessment of ovarian reserve testing prior to IVF (Hendriks *et al.*, 2005b).

Materials and Methods

Sources

Ovarian reserve tests that were considered as potential part of a predictive model were basal FSH, estradiol, inhibin B and AMH, as well as sonographic markers such as AFC, total OV and ovarian blood flow. Moreover, female age, being an indicator for ovarian reserve, was also considered as potential prognosticator. Challenge tests like the CCCT or GAST were also included in the search.

For every ORT a computerized MEDLINE search was performed to identify articles on the subject published from 1978 until April 2006. Keywords used for the various searches were 'in vitro fertilisation' or 'in vitro fertilization' or 'assisted' or 'intracytoplasmic' or 'intracytoplasmatic', in combination with 'test-specific' keywords as published elsewhere (Broekmans *et al.*, 2006). To be selected for further reading abstracts that were recognised by the search had to report clearly on one or two outcomes of IVF or ICSI treatment: poor response and/or pregnancy.

Selected studies were carefully read and scored by two of the authors (D.J.H and F.J.M.B) and judged upon the presence of multivariate prediction models that enabled the construction of 2x2 tables to calculate test characteristics. The authors of studies containing multivariate models in which it was not possible to construct 2x2 tables were asked to supply us with data needed for the construction of 2x2 tables or otherwise these studies were excluded. In addition, cross-references of the selected studies were checked for other articles meeting the inclusion criteria. All articles, selected according to the previously mentioned search strategy were checked using multivariate prediction models.

Two-by-two tables comparing results of the multivariate prediction models and the

occurrence of poor ovarian response and/or pregnancy were constructed independently by two of the authors (D.J.H and F.J.M.B) and in the event of disagreement, the judgement of a third author (B.W.J.M) was decisive. For each study, we calculated the prevalence of poor ovarian response and/or pregnancy. There were no uniform criteria for the definition of poor ovarian response. In this analysis poor response encompassed cycle cancellation as well as insufficient follicular growth or oocyte yields according to the standards of each study. Also, data on clinical and ongoing pregnancies were not analysed separately. The following characteristics of each study were registered: (1) sampling (consecutive versus other), (2) data collection (prospective versus retrospective), (3) study design (cohort study versus case-control study), (4) blinding (present or absent), (5) selection bias and (6) verification bias (Lijmer *et al.*, 1999). The definition of poor response or pregnancy was documented as well as whether an included study reported on only one cycle per couple or on multiple cycles.

Analysis

The analysis was conducted according to a methodology that has been described in detail elsewhere (Bancsi *et al.*, 2003). First, we assessed the predictive performance of each model in detail. In brief, for each study sensitivity and specificity were calculated from the published data. Subsequently, sensitivity-specificity points for each study were plotted in ROC space. Homogeneity of the studies was tested by means of the chi-square test. A summary point estimate of sensitivity and specificity was calculated if homogeneity could not be rejected. In case of heterogeneity, logistic regression was used to evaluate whether the study characteristics influenced the discriminative capacity of the models. A P-value <0.05 was considered to indicate statistical significance. If one of the study characteristics was found to have a statistically significant impact on the performance of the model, further analysis was performed in subgroups of patients.

A Spearman correlation coefficient was calculated for the association between sensitivity and specificity to explore possible heterogeneity due to a shift in cut-off levels between the models. If there was a negative correlation between sensitivity and specificity as defined by a correlation coefficient of -0.5 or less, a summary ROC-curve was estimated (Littenberg and Moses, 1993; Moses *et al.*, 1993; Midgette *et al.*, 1993), using a random-effects regression model (SAS Institute Inc., 1993).

To compare the predictive capacity of the multivariate models with that of AFC, the constructed summary ROC-curves for the multivariate models and a previously constructed curve of AFC were tested for statistical significant differences using linear regression (Hendriks *et al.*, 2005b).

Results

Search Strategy

A total of 11 studies reporting on the predictive capacity of several multivariate models were identified and considered suitable for data extraction and meta-analysis (Balasch *et al.*, 1996; Ranieri *et al.*, 1998; Creus *et al.*, 2000; Fabregues *et al.*, 2000; Bancsi *et al.*, 2002; van Rooij *et al.*, 2003; Durmusoglu *et al.*, 2004; Erdhem *et al.*, 2004; Muttukrishna *et al.*, 2004; Hendriks *et al.*, 2005a; Penarrubia *et al.*, 2005). All studies reported on poor ovarian response as outcome, but only one of the studies reported on pregnancy as outcome (Creus *et al.*, 2000).

Characteristics of the included studies are listed in Table 1. Sampling of patients was consecutive in eight studies and data collection was prospective in five studies. Seven studies were designed as cohort studies. None of the studies was blinded. Selection bias was present in six of 11 included studies, whereas verification bias was present in two studies. All studies reported on the outcome of one cycle per couple.

As with most studies on ORT, definitions of poor response varied considerably. Each of the 11 studies used a different model. All studies reported on models that were constructed by the authors themselves in a so-called internal validation, whereas studies validating an existing prediction model were not present. The number of variables combined in the different models varied from two to four. The prediction variables used were age, inhibin B, FSH, AFC, AMH, GAST, OV and the CCCT. In the models, FSH was used seven times,

TABLE 1. Characteristics of the 11 included studies on multivariate models on ovarian reserve testing

Author	Consecutive	One cycle per couple	Data per cycle	Cohort / Case-control	Prospective / Retrospective	Blinding	Selection bias	Verification bias
Balasch <i>et al.</i> (1996)	no	yes	yes	case-control	retrospective	no	no	no
Ranieri <i>et al.</i> (1998)	no	yes	yes	cohort	prospective	no	no	no
Creus <i>et al.</i> (2000)	yes	yes	yes	case-control	retrospective	no	yes	no
Fabregues <i>et al.</i> (2000)	yes	yes	yes	case-control	retrospective	no	yes	no
Bancsi <i>et al.</i> (2002)	yes	yes	yes	cohort	prospective	no	no	no
Van Rooij <i>et al.</i> (2002)	yes	yes	yes	cohort	prospective	no	no	no
Durmusoglu <i>et al.</i> (2004)	yes	yes	yes	cohort	retrospective	no	yes	no
Erdhem <i>et al.</i> (2004)	yes	yes	yes	cohort	retrospective	no	yes	yes
Muttukrishna <i>et al.</i> (2004)	no	yes	yes	cohort	prospective	no	yes	no
Hendriks <i>et al.</i> (2005)	yes	yes	yes	cohort	prospective	no	no	no
Penarrubia <i>et al.</i> (2005)	yes	yes	yes	case-control	retrospective	no	yes	yes

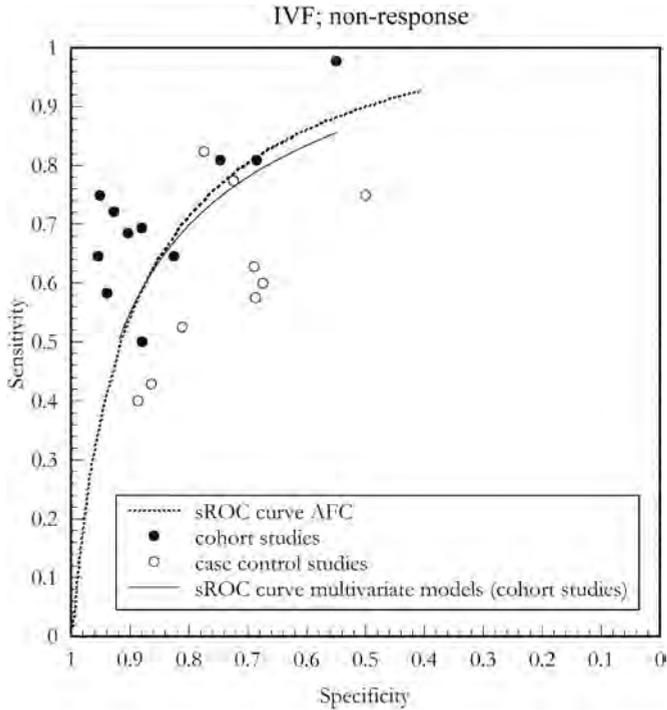


FIGURE 1. Receiver operating characteristic space with the reported sensitivities and specificities of the multivariate models. The closed circles represent the estimates from cohort studies whereas the open circles represent the estimates from case-control studies. Furthermore, three summary ROC-curves are shown.

inhibin B six times, age five times, AFC four times, CCCT and AMH two times and GAST or OV each only one time. The combinations of prediction variables that were most frequent were FSH + inhibin B (Balasch *et al.*, 1996; Creus *et al.*, 2000; Fabregues *et al.*, 2000; Bancsi *et al.*, 2002) and age + FSH (Balasch *et al.*, 1996; Creus *et al.*, 2000; Durmusoglu *et al.*, 2004; Penarrubia *et al.* 2005).

The sensitivities and specificities for the prediction of poor ovarian response, as calculated from each model, are summarised in Table 2. The sensitivity varied between 39% and 97%, and the specificity between 50% and 96%. A plot of sensitivity-specificity points in a ROC-space is shown in the Figure 1.

Logistic regression analysis indicated that cohort studies reported a significantly better discriminative performance than case-control studies, whereas none of the other study characteristics had a statistically significant impact on the reported performance of the different models. As a consequence, further analysis was performed separately in the cohort studies and case-control studies.

TABLE 2. Prediction variables used in different models to predict ovarian reserve in IVF patients

Author	Prediction variables	Poor response (n)	Normal response (n)	Total (n)	Sensitivity	Specificity
Balasch <i>et al.</i> (1996)	Age + FSH	40	80	120	53	81
	Age + inhibin B	40	80	120	59	67
	Inhibin B + FSH	40	80	120	57	69
	Age + FSH + inhibin B	40	80	120	39	89
Ranieri <i>et al.</i> (1998)	FSH + GAST	48	129	177	97	55
Creus <i>et al.</i> (2000)	Age + FSH	40	80	120	83	77
	Age + inhibin B	40	80	120	74	50
	FSH + inhibin B	40	80	120	77	73
	Age + FSH + inhibin B	40	80	120	83	77
Fabregues <i>et al.</i> (2000)	FSH + inhibin B	28	52	80	42	86
Bancsi <i>et al.</i> (2002)	AFC + inhibin B + FSH	36	84	120	75	95
	AFC + FSH	36	84	120	72	93
	AFC + inhibin B	36	84	120	69	88
	FSH + inhibin B	36	84	120	58	94
Van Rooij <i>et al.</i> (2002)	AMH + inhibin B + FSH	35	84	119	68,6	90,5
Durmusoglu <i>et al.</i> (2004)	Age + AFC	24	67	91	52	88
Erdem <i>et al.</i> (2004)	Abnormal CCT + age	16	16	32	81	69
	Abnormal CCT + age + OV + AFC	16	16	32	81	75
Muttukrishna <i>et al.</i> (2004)	FSH + inhibin B + AMH§	17	52	69	62,5	83,3
Hendriks <i>et al.</i> (2005)	FSH + AFC	17	46	63	65	96
Penarrubia <i>et al.</i> (2005)	Age + FSH	129	129	258	63,1	68,8

CCT=clomiphene citrate challenge test; OV=ovarian volume; AFC=antral follicle count; FSH=follicle stimulating hormone; GAST gonadotrophin agonist stimulation test; AMH=anti-mullerian hormone

§ Threshold values: FSH > 10 mIU/ml, inhibin B > 80 pg/ml and AMH > 0,1 ng/ml

In the seven cohort studies, homogeneity was rejected for both sensitivity and specificity (both P-values < 0.001), and therefore a summary point estimate was not given. The Spearman correlation coefficient between sensitivity and specificity was -0.77 (P-value 0.006). This indicated that an increase in sensitivity occurred when specificity dropped, or vice versa. Thus, the combination of sensitivity-specificity points could be thought of as originating from an ROC curve, in which variations in sensitivity-specificity were explained by different cut-off values. Therefore, we summarized the test-performance of the multivariate models by estimating a summary ROC curve (Fig. 1). This summary ROC curve had a shape almost similar to the summary ROC curve for the AFC as published elsewhere (Hendriks *et al.*, 2005b).

In the four case-control studies, homogeneity was also rejected for both sensitivity and specificity (both P-values < 0.001). The Spearman correlation coefficient for sensitivity and

specificity was -0.50 (P-value 0.006). As cohort studies are methodologically superior to case-control studies, a separate analysis of the case-control studies was not performed. Figure 1 also shows the summary ROC curves of the AFC that was constructed in a previous meta-analysis (Hendriks *et al.*, 2005b). Comparison of this summary ROC curve of AFC with the summary ROC curve of the multivariate models as estimated from cohort studies did not indicate a better performance of multivariate models than AFC (P-value 0.45). Only one study reported results of the prediction of non-pregnancy (Creus *et al.*, 2000). The variables analysed were age, basal FSH and basal inhibin B. The likelihood of pregnancy was analysed and age had the highest area under curve in the ROC curve (0.72), which was better than basal FSH or basal inhibin B alone (both 0.60) or any combination of the three variables.

Discussion

The results of this meta-analysis show that multivariate models as a test for the occurrence of poor ovarian response after IVF have a similar potential compared to AFC alone. However, no data are available on the predictive capacity of multivariate models for pregnancy, let alone live birth. According to these findings, there is at present no place for multivariable models in the prediction of ovarian reserve prior to IVF. Future models, should take into account the AFC, as this test is at present the best univariate predictor of diminished ovarian reserve.

A concern in the meta-analysis of diagnostic studies is heterogeneity among the included studies with respect to the definitions used, study design, and clinical characteristics of the population studied (Irwig *et al.*, 1994; Deeks, 2001). We found a clear difference between cohort studies and case-control studies. Surprisingly, cohort studies reported a better performance than case-control studies. Other sources of heterogeneity might be caused by differences in patient population, different treatment protocols used, and variability in the definition of poor ovarian response. Despite this heterogeneity, we were able to provide an estimate summary ROC curve, which we feel is at present the best available evidence on the subject. Especially our finding that multivariate models did not perform better than AFC alone is in our opinion of clinical importance.

This meta-analysis indicates that AFC performed equally to multivariate models in assessing quantitative ovarian reserve. However, we are aware of the fact that we

used different populations in the comparison of both tests in the present study. This could hamper comparability of the performance of these tests. It would be possible to overcome this problem by limiting the analysis to studies reporting on the predictive capacity of both test approaches in the same group of women. Only three studies in this meta-analysis incorporated AFC in their models (Bancsi *et al.*, 2002; Durmusoglu *et al.*, 2004, Hendriks *et al.*, 2005a). Durmusoglu *et al.* (2004) found an area under the ROC curve for basal AFC of 0.82 with an optimum cut-off for the prediction of poor response of six primordial follicles. Sensitivity and specificity for the AFC were 85% and 74%, respectively. This performance was not improved by adding female age together with the AFC in a predictive model. Bancsi *et al.* (2002) evaluated inhibin B, basal FSH and AFC. AFC as single variable was the best predictor of poor response with a sensitivity of 61% and specificity of 88%. The area under the ROC curve for AFC was 0.87, better than for inhibin B (0.77) and FSH (0.84). All the multivariate models constructed in this study had a better overall ability to discriminate between normal and poor response, with a maximum increase of the ROC-AUC to a value of 0.92 for a model containing AFC, basal FSH and inhibin B. With this model a sensitivity of 0.75 and specificity of 0.95 was obtained (see Table 2). In the study from Hendriks *et al.* (2005b) the best single variable predictor of poor response was basal FSH with an area under the ROC curve of 0.85 followed by AFC with the area under the ROC curve of 0.83. Combining these two variables yielded the best predictor model for ovarian reserve (AUC-ROC 0.89), but with a modest improvement of sensitivity and specificity. All this indicates that multifactor models may have some benefit over using the AFC as a single test, although the added clinical value has not been consistently shown.

Recent years have shown several studies on the role of AMH as univariate predictor of outcome after IVF. AMH is likely to represent the size of the primordial follicle pool as accurately as the AFC, as small antral follicles produce AMH, and AFC relate closely to primordial follicle numbers (Kevenaar *et al.*, 2006). Also, antral follicle numbers measured on day 3 are highly correlated to the serum levels of AMH (Fanchin *et al.*, 2003). From several studies it has been shown that AMH levels do not fluctuate across and between cycles making it a cycle independent marker (Hehenkamp *et al.*, 2006; La Marca *et al.*, 2006; Tsepelidis *et al.*, 2007). In contrast, the AFC is more likely to be a marker that is more prone to observer bias and at least varies more clearly between cycles in patients (Fanchin *et al.*, 2003; van Rooij *et al.*, 2005). All this may allow AMH to become an easy and powerful tool for quantitative ovarian reserve testing. Several clinical studies suggest AMH to have the same accuracy compared to the AFC in predicting the poor responder

after IVF, although pregnancy prediction continues to be troublesome (van Rooij *et al.*, 2005; Elgindy *et al.*, 2007; Kwee *et al.*, 2007; La Marca *et al.*, 2007; McIlveen *et al.*, 2007). Meta-analysis of the accumulating studies on AMH may further exemplify its merits in this field compared to the AFC.

The performance of a prognostic model can be assessed in the sample or population in which it has been developed (apparent or internal validity) or in other populations, i.e. patients seen in another period or at another place (external validity or generalizability). All studies included in the present meta-analysis reported on models that were assessed in the sample or population in which they had been developed. External validation is needed to assess the true accuracy of the models that we identified (Mol *et al.*, 2000). Such validation studies are lacking for ORT. However, most studies on prediction models in reproductive medicine have shown that the internal model often overestimates the true predictive capacity (Stolwijk *et al.*, 1996).

Even if a test has good predictive accuracy, this does not mean that the test is good enough to have clinical implications. For assessment of the clinical applicability of a test, the relative weight of both false positive and false negative predictions should be considered. When patients are interviewed on issues of incorrect withholding IVF as compared to incorrect performing IVF, they value the first much worse than the second, thereby implicating that the present tests for ovarian reserve have in itself insufficient accuracy to withhold IVF (Mol *et al.*, 2006). As a consequence from this point of view, testing for ovarian reserve is not needed. Whether incorporation of the AFC in future models improves the predictive capacity should be answered in future studies.

Apart from the prediction of pregnancy or live birth after IVF, the prediction of ovarian reserve is also potentially important for individual adjustment of the dose of gonadotropins prior to IVF. Patients with an expected poor response might benefit from a higher starting dose than 150 IU/day. A randomised trial on the subject showed that an individual dose regimen in a well-defined 'standard' patient population increased the proportion of appropriate ovarian responses and decreased the need for dose adjustments during controlled ovarian stimulation (Popovic-Todorovic *et al.*, 2003). A higher ongoing pregnancy rate was observed in the individual dose group. In contrast, no beneficial effect was observed in a randomised dose study of predicted poor responders (Klinkert *et al.*, 2005). This all implies that additional studies are needed to provide clear evidence for the beneficial effects of dose adaptation in poor responder patients, as many

previous studies have failed to show such effect (Tarlatis *et al.*, 2006). In our opinion, such studies provide more information, when the randomization is limited to women in whom the prediction model indicates a starting dose different than the standard dose of 150 IU/day (Bossuyt *et al.*, 2000). Furthermore the endpoint of such a study should be live birth after IVF and not only the ovarian response to stimulation (Farquhar, 2000).

In conclusion, this study shows that the performance of multivariate models in the prediction of poor ovarian response after IVF is comparable to that of AFC. Therefore, the AFC may be considered as the test of first choice in the assessment of diminished ovarian reserve. Future models for the prediction of poor response and pregnancy should incorporate the AFC.

References

- Abma JC, Chandra A, Mosher WD, Peterson CS and Piccinino CJ. Fertility, family planning and women's health: new data from the 1995 national survey of family growth. *Vital Health Stat* 1997; 23:1-114.
- Balasch J, Creus M, Fabregues F, Carmona F, Casamitjana R, Ascaso C, Vanrell JA. Inhibin, follicle-stimulating hormone, and age as predictors of ovarian response in in vitro fertilization cycles stimulated with gonadotropin-releasing hormone agonist-gonadotropin treatment. *Am J Obstet Gynecol* 1996;175:1226-1230.
- Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD and te Velde ER. Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. *Fertil Steril* 2002;77:328-336.
- Bancsi LF, Broekmans FJ, Mol BW, Habbema JD and te Velde ER. Performance of basal follicle-stimulating hormone in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. *Fertil Steril* 2003;79:1091-100.
- Bossuyt PM, Lijmer JG and Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;356:1844-7.
- Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update* 2006;12:685-718.
- Creus M, Penarrubia J, Fábregues F, Vidal E, Carmona F, Casamitjana R, Vanrell JA and Balasch J. Day 3 serum inhibin B and FSH and age as predictors of assisted reproduction treatment outcome. *Hum Reprod* 2000;15:2341-2346.
- Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157-62.
- Durmusoglu F, Elter K, Yoruk P and Erenus M. Combining day 7 follicle count with the basal antral follicle count improves the prediction of ovarian response. *Fertil Steril* 2004;81:1073-1078.
- Elgindy EA, El-Haieg DO, El-Sebaey A. Anti-Müllerian hormone: correlation of early follicular, ovulatory and midluteal levels with ovarian response and cycle outcome in intracytoplasmic sperm injection patients. *Fertil Steril* 2008;89:1670-1676.
- Erdhem M, Erdhem E, Gursoy R and biberoglu K. Comparison of basal and clomiphene citrate induced FSH and inhibin B, ovarian volume and antral follicle counts as ovarian reserve tests and predictors of poor ovarian response in IVF. *J Assist Reprod Genet* 2004;21:37-45.
- Fábregues F, Balasch J, Creus M, Carmona F, Puerto B, Quinto L, Casamitjana R and Vanrell JA. Ovarian reserve test with human menopausal gonadotropin as a predictor of in vitro fertilization outcome. *J Assist Reprod Genet* 2000;17:13-19.
- Fanchin R, Schonäuer LM, Righini C, Frydman N, Frydman R, Taieb J. Serum anti-Müllerian hormone dynamics during controlled ovarian hyperstimulation. *Hum Reprod* 2003;18:328-32.
- Farquhar CM. Extracts from the 'clinical evidence'. *Endometriosis. BMJ* 2000;321:1077-8.

Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, Te Velde ER, Broekmans FJ. Anti-Müllerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. *J Clin Endocrinol Metab* 2006;91:4057-63.

Hendriks DJ, Broekmans FJ, Bancsi LF, de Jong FH, Looman CW and te Velde ER. Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve. *Hum Reprod* 2005a;20:163-169.

Hendriks DJ, Mol BW, Bancsi LF, Te Velde ER, Broekmans FJ. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. *Fertil Steril* 2005b;83:291-301.

Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;120:667-76.

Kevenaar ME, Meerasahib MF, Kramer P, van de Lang-Born BM, de Jong FH, Groome

NP, Themmen AP, Visser JA. Serum anti-mullerian hormone levels reflect the size of the primordial follicle pool in mice. *Endocrinology* 2006;147:3228-34.

Klinkert ER, Broekmans FJ, Looman CW, Te Velde ER. A poor response in the first in vitro fertilization cycle is not necessarily related to a poor prognosis in subsequent cycles. *Fertil Steril* 2004;81:1247-53.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. *Hum Reprod* 2005;20:611-5.

Kwee J, Schats R, McDonnell J, Themmen A, de Jong F, Lambalk C. Evaluation of anti-Müllerian hormone as a test for the prediction of ovarian reserve. *Fertil Steril* 2008;90:737-743.

La Marca A, Stabile G, Artesio AC, Volpe A. Serum anti-Mullerian hormone throughout the human menstrual cycle. *Hum Reprod* 2006;21:3103-7

La Marca A, Giulini S, Tirelli A, Bertucci E, Marsella T, Xella S, Volpe A. Anti-Müllerian hormone measurement on any day of the menstrual cycle strongly predicts ovarian response in assisted reproductive technology. *Hum Reprod* 2007;22:766-71.

Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making* 1993;13:313-21.

Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH and Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests.

JAMA 1999;282:1061-6.

Midgette AS, Stukel TA and Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Med Decis Making* 1993;13:253-7.

Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993;12:1293-316.

Mol BW, Verhagen TE, Hendriks DJ, Collins JA, Coomarasamy A, Opmeer BC and Broekmans FJ. Value of ovarian reserve testing before IVF: a clinical decision analysis. *Hum Reprod* 2006;21:816-23.

Mol BW, van Wely M and Steyerberg EW. Using prognostic models in clinical infertility. *Hum Fertil (Camb)* 2003;3:199-202.

Muttukrishna S, Suharjono H, McCarrigle H and Sathanandan M. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? *BJOG* 2004;111:1248-1253.

Muttukrishna S, McCarrigle H, Wakim R, Khadum I, Ranieri DM and Serhal P. Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? *BJOG* 2005;112:1384-1390.

Penarrubia J, Fabregues F, Manau D, Creus M, Carmona F, Casamitjana R, Vanrell J and Balasch J. Previous cycle cancellation due to poor follicular development as a predictor of ovarian response in cycles stimulated with gonadotropin-releasing hormone agonist-gonadotropin treatment. *Hum Reprod* 2005;20:622-628.

Popovic-Todorovic B, Loft A, Bredkjaer HE, Bangsbo S, Nielsen IK and Andersen AN. A prospective randomized clinical trial comparing an individual dose of recombinant FSH based on predictive factors versus a 'standard' dose of 150 IU/day in 'standard' patients undergoing IVF/ICSI treatment. *Hum Reprod* 2003;18:2275-82.

Quban HS, Malkawi HY, Tahat YA, Areidah S, Nusair B, Khreisat BM, Al-Quraan G, Abu-Assaf A, Hadaddein MF and Abu-Jassar H. In-Vitro fertilisation treatment: factors affecting its results and outcome. *J Obstet Gynaecol* 2005;25:689-93.

Ranieri DM, Quinn F, Makhlof A, Khadum I, Ghutmi W, McGarrigle H, Davies M and Serhal P. Simultaneous evaluation of basal follicle-stimulating hormone and 17 β -estradiol response to gonadotropin-releasing hormone analogue stimulation: an improved predictor of ovarian reserve. *Fertil Steril* 1998;70:227-233.

SAS Institute Inc. SAS Technical Report P-242, SASR: Changes and Enhancements, Release 6.08. Cary NC: SAS Institute INC, 1993;113-116.

Stolwijk AM, Hamilton CJCM, Hollanders JMG, Bastiaans LA, Zielhuis GA. A more realistic approach to the cumulative pregnancy rate after in-vitro fertilization. *Hum Reprod* 1996;11:660-663.

Tarlatzis BC, Fauser BC, Kolibianakis EM, Diedrich K, Rombauts L, Devroey P. GnRH antagonists in ovarian stimulation for IVF. *Hum Reprod Update* 2006;12:333-40.

Templeton A, Morris JK and Parslow W. Factors that affect outcome of in-vitro fertilisation treatment [see comments]. *Lancet* 1996;348:1402-1406.

Tsepelidis S, Devreker F, Demeestere I, Flahaut A, Gervy Ch, Englert Y. Stable serum levels of anti-Müllerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. *Hum Reprod* 2007;22:1837-40.

Te Velde ER and Pearson PL. The variability of female reproductive aging. *Hum Reprod Update* 2002;8:141-154.

Van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH and Themmen AP. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. *Hum Reprod* 2002;17:3065-3071.

Van Rooij IA, Broekmans FJ, Scheffer GJ, Looman CW, Habbema JD, de Jong FH, Fauser BJ, Themmen AP, te Velde ER. Serum antimullerian hormone levels best reflect the reproductive decline with age in normal women with proven fertility: a longitudinal study. *Fertil Steril* 2005;83:979-87.

Weinstein M, Wood JW and Chang MC. Age patterns in fecundability. In Gray R, Leridon H and Spina A (eds) *Biochemical and demographic determinants of reproduction*. Oxford: Clarendon Press, 1993:209-220.

Chapter 8

Summary and discussion
Samenvatting

Summary

The definition of subfertility is not having been able to conceive after at least 12 months of unprotected intercourse. It affects approximately 9-15% of couples of reproductive age (Beurskens *et al.*, 1995; Boivin *et al.*, 2007). The incidence of subfertility is increasing in the developed world mainly due to postponed maternity. A subfertile couple can be referred to the hospital for a diagnostic fertility work-up. Standard investigations, including semen analysis, tubal patency tests and assessment of ovulation will be performed. This work-up has two aims: find an explanation for the subfertility and calculate the probability of a spontaneous pregnancy in the next year using a prognostic model (NVOG-guideline, 2004). After the fertility work-up about 20% of couples is diagnosed with ovulation disorders, 25% with male subfertility, 25% with tubal obstruction or endometriosis and 30% with unexplained subfertility (Collins and Van Steirteghem, 2004).

When a cause for the couple's subfertility is found, for example anovulation, blocked tubes or a severe sperm defect, an appropriate treatment will be suggested. In couples with unexplained or mild male subfertility (about 50% of couples), expectant management, intra-uterine insemination (IUI), with or without ovarian hyperstimulation, or in vitro fertilization (IVF) are possible strategies. But IUI or IVF will not always result in higher pregnancy or live birth rates than expectant management, and are expensive options with risks of adverse events (multiple pregnancies or ovarian hyperstimulation syndrome (OHSS)) (Bensdorp *et al.* 2007; Pandian *et al.* 2012; Veltman-Verhulst *et al.* 2012).

To be able to counsel subfertile couples adequately, we have to know what their motives and goals are when they decide to visit a physician for help. And we have to know the success and dropout rates of fertility treatment as well as the behaviour of treated couples during fertility treatment. This thesis addresses several parts of the fertility treatment subfertile couples will experience, when seeking advice and treatment for their fertility problem.

Chapter 1 gives an outline and describes the objectives of this thesis.

In **chapter 2** the results of questionnaires that were administered to 38 subfertile couples (38 women and 28 of their male partners) referred to the Máxima Medical Centre for the diagnostic fertility work-up were presented. Finding a reason why they had not conceived

was a more important motivation for participants to enroll in the fertility work-up than getting pregnant. Informed shared decision making on the fertility work-up was preferred by 95%. The majority of participants wanted to know as much as possible about the diagnostic fertility work-up (73%), fertility treatment (70%) and decreased fertility (62%). Women reported a higher level of knowledge about the fertility work-up than men. In the virtual trade-off for participants who initially had chosen the diagnostic fertility work-up, the mean rate at which their preference switched from fertility work-up to six months expectant management was with a risk of a complication of 20% during the work-up (23% for women and 15% for men; $p=0.22$). Half of participants preferred an expectant management when the risk of a complication during the work-up exceeded 10%. The mean effectiveness rate at which patients switched preference was 24% (26% for women and 21% for men, $p=0.35$). The subfertile couples interviewed in this study preferred informed and shared decision making and they mainly want to find an explanation for not getting pregnant during fertility work-up. This information can be used for counseling couples referred for the diagnostic fertility work-up.

In **chapter 3** the incidence of couples dropping out of the in vitro fertilization (IVF) waiting list was determined and the couples' reasons for dropping out described. Data from 674 women consecutively placed on the IVF waiting list in Maastricht University Medical Centre between June 2000 and July 2003 were retrieved. Follow-up information was collected in 2005 and 2008. Of the 674 couples on the waiting list, 87% started IVF, and 13% dropped out before starting their first IVF cycle. Follow-up data were obtained for 85/86 patients (98.8%): 32 (37%) dropped out because of spontaneous pregnancy, 31 (36%) for personal reasons (passive censoring), and 23 (27%) for medical reasons (active censoring). Most pregnancies occurred within 3 months after the patient had been placed on the waiting list (30 of 32, 94%). Of the 54 censored couples, four became pregnant. We concluded that on a 6-months waiting list for IVF, 13% of the couples dropped out before starting treatment. The most important reason for dropout was (spontaneous) pregnancy. Most of these pregnancies occurred within 3 months, which suggests that psychological factors such as stress relief after being placed on the waiting list might be operative.

In **chapter 4** the impact of dropout in IVF-programmes on the reported cumulative pregnancy rate was assessed with follow-up data retrieved from 588 couples starting their first IVF or ICSI-treatment in the Maastricht University Medical Centre between June 2000 and July 2003. Data from all 588 couples were collected retrospectively (the 'as treated group'). In 91% of dropouts, follow-up data could be obtained by returned questionnaires

from the patients or information from their general practitioner. The data from the treated group supplemented with the achieved follow-up data from the dropouts, defined the 'completed group'. Cycle-based and real-time-based cumulative pregnancy rates (CPRs) were calculated for the 'as treated group' using three assumptions for dropouts: (I) dropouts having no probability of pregnancy, (II) dropouts having the same probability of pregnancy as those continuing treatment and (III) dropouts discontinuing because of medical reasons having no chance of pregnancy and dropouts discontinuing because of other reasons having the same probability of pregnancy as those continuing treatment. CPRs obtained in the 'as treated group' were compared to CPRs calculated using the data set including the follow-up data of the dropouts ('completed group'). The cycle-based CPR after three IVF-cycles ranged from 63-71% in the 'as treated group' and was 65% in the 'completed group'. The real-time-based CPR after nine months ranged from 54-59% in the 'as treated group' and was 55% in the 'completed group'. The pregnancy rate (PR) in dropouts was 14% (95% confidence interval (CI) 8, 22%). We concluded that in IVF programmes, where outcome data of dropouts remain unknown, CPRs should be calculated by assuming dropouts to have a PR between no probability and the same probability as those who continue treatment. Our study showed that the most accurate estimate for the PR in dropouts is 14%.

In **chapter 5** the effect of dropout on pregnancy rates in the next treatment cycle was evaluated. It has been suggested that pregnancy rates reported in some fertility treatment studies are inflated due to high dropout rates, but empirical data are lacking. A literature search was performed for studies reporting on patients undergoing multiple cycles of IUI or IVF/ICSI. Studies were included if they allowed extraction of dropout and pregnancy rates per cycle. For each cycle we calculated the dropout rate and the pregnancy rate. Data on IUI and IVF were analysed separately. We used Pearson correlation coefficient to evaluate whether dropout rate was associated with pregnancy rate. Overall, we analysed data of 17,555 couples, who underwent 52,003 cycles of IUI and of 45,934 couples, who underwent 99,345 cycles of IVF. For IUI, pregnancy rates ranged from 0% to 50% per cycle (mean 8%; 95% CI 7, 9) and dropout rates ranged from 4% to 85% per cycle (mean 30%; 95% CI 27, 33). For IVF, pregnancy rates ranged from 7% to 36% per cycle (mean 18%; 95% CI 16, 19), and dropout rates ranged from 7% to 69% per cycle (mean 36%; 95% CI 32, 41). Neither for IUI nor for IVF we found a significant relation between pregnancy rate and dropout rate in the preceding cycle. Overestimation of pregnancy rates in ART due to dropout seemed limited.

In **chapter 6** the value of testing for ovarian reserve prior to a first cycle of IVF was assessed. A decision model was developed for couples considering participation in an IVF program. Three strategies were evaluated: (I) withhold IVF without prior ovarian reserve testing, (II) test for ovarian reserve, and start IVF treatment if ovarian reserve was estimated to be sufficient, and (III) start IVF without prior ovarian reserve testing and withhold further treatment in case of poor response. The outcome considered was the birth of a child. The valuation of the combination of the strategy conducted and the outcome accomplished was expressed on a distress scale in units of 'IVF cycles that were performed in vain'. Correct treatment with IVF and correct withholding of IVF were considered to bring no distress. The distress of withholding IVF in case pregnancy would have occurred is consequently specified by the ratio of the expected distress after incorrectly withholding IVF to the expected distress after incorrectly performing IVF (distress ratio). We interviewed both patients and physicians to determine realistic estimates for this distress ratio. The value of testing for ovarian reserve strongly depends on the expected pregnancy rate after IVF as well as on the valuation of the incorrect decisions from testing. For realistic ranges of the success rate after IVF and for distress ranges as were measured, treatment of all couples without testing was found to generate less distress than testing for ovarian reserve. The sensitivity and specificity of testing for ovarian reserve has to improve to 50% and 96% respectively, to make testing a valuable strategy. Based on the decision analysis, where current test accuracy and preference inventory among patients and physicians were used, testing for ovarian reserve seemed not useful.

In **chapter 7** we performed a meta-analysis to review the accuracy of multivariate models to predict ovarian reserve and pregnancy in women undergoing IVF compared with the antral follicle count (AFC) as a single test. A MEDLINE and EMBASE search was performed to identify articles published on multivariable models for ovarian reserve testing in patients undergoing IVF. For the selected studies, sensitivity and specificity of the test in the prediction of poor ovarian response and of non-pregnancy were calculated. Overall performance was assessed by estimating a summary Receiver Operating Characteristic (ROC) curve, which was compared to the current best single test, the ROC curve for AFC. We identified 11 studies reporting on the predictive capacity of multivariate models in ovarian reserve testing. All studies reported on the prediction of poor ovarian response, whereas none reported on the occurrence of pregnancy. The sensitivity for prediction of poor ovarian response varied between 39% and 97% and the specificity between 50% and 96%. Logistic regression analysis indicated that cohort studies provided a

significantly better discriminative performance than case-control studies. As cohort studies are superior to case-control studies further analysis was limited to the cohort studies. For the cohort studies, a summary ROC curve could be estimated, which had a similar shape as a previously constructed summary ROC curve for the AFC by Hendriks *et al.* (2005). We concluded that the accuracy of multivariate models for the prediction of ovarian response in women undergoing IVF is similar to the accuracy of AFC. No data are available on the capacity of these models to predict pregnancy, let alone live birth. On the basis of these findings, the use of more than one single test for the assessment of ovarian reserve could not be supported.

General discussion

In fertility care it is of the utmost importance to find out which subfertile couple needs which treatment at what moment. The diagnostic fertility work-up gives us the information to differentiate between couples who require assisted reproduction treatment and couples who still have a good chance of a spontaneous conception. The results of the questionnaires we administered to couples before the diagnostic fertility work-up was started showed that they were more interested in the explanation why they do not achieve a pregnancy than in getting pregnant as such (chapter 2). This is one explanation why almost half of subfertile couples referred to the hospital discontinue before any stage of fertility treatment was started; the main reasons to discontinue were 'reject fertility treatment in general' and 'relational problems' (Brandes *et al.* 2009).

Ideally we do not treat subfertile couples with a good chance of a spontaneous pregnancy as long as their natural chance exceeds their chance of a pregnancy due to ART. ART is an expensive treatment and carries risks, e.g. multiple pregnancies and ovarian hyperstimulation syndrome. On the other hand, couples with no chance of a natural conception (infertile couples) have to be identified so they can be offered ART after being counselled. Our follow-up reports show that spontaneous pregnancies occur on an IVF waiting list and is the main reason for discontinuing treatment at that stage (chapter 3). 14% of couples discontinuing IVF treatment reported a spontaneous pregnancy (chapter 4). Do these percentages mean that we place couples on a waiting list for IVF too soon? Do we doctors have to observe more patience and motivate couples to take more time to get pregnant spontaneously? The strategy of expectant management in subfertile couples with a good/ intermediate prognosis of spontaneous pregnancy is called tailored

expectant management (TEM). Optimal adherence to TEM can lead to the prevention of unnecessary treatment, complications, costs and physical and psychological burdens. Van den Boogaard evaluated professionals 'and patients' barriers and facilitators for tailored expectant management (Van den Boogaard *et al.*, 2012). From these studies can be concluded that the implementation of tailored expectant management in couples with a good prognosis can be enhanced by developing adequate patient information material, implementing regular fertility meetings where patients are discussed after the diagnostic fertility work-up, implementing local protocols and by improving the knowledge of doctors about prognostic models and their communication skills with interactive training sessions (Van den Boogaard *et al.* 2011; Van den Boogaard *et al.* 2012).

Couples discontinue fertility treatment for many reasons. These can roughly be divided in two categories: medical reasons (active censoring) and personal reasons (passive censoring) (Land *et al.* 1997; Smeenk *et al.* 2003). One of the medical reasons mentioned before is a poor prognosis due to poor response. Personal reasons are mainly emotional distress, relation problems and financial problems when ART is not reimbursed (Goldfarb *et al.* 1997; Domar 2004; Brandes *et al.* 2009). When counselling couples for fertility treatment information about pregnancy rates and dropout rates should be provided. The pregnancy rate per treatment cycle is easy to calculate but has limited value for individuals because it does not account for the potential need for multiple treatment cycles. Longitudinal analysis of live birth rates instead of cross-sectional reports provides a better outcome and gives individuals a more realistic prognosis of ART (Malizia *et al.*, 2009). In ART programmes, outcome data of couples discontinuing treatment remain unknown. In our follow-up study we found that cumulative pregnancy rates should be calculated by assuming dropouts to have a PR between no probability and the same probability of a pregnancy as those who continue treatment (chapter 4). Pregnancy rates are decreasing in consecutive IUI and IVF cycles (chapter 5) and this information is also important in counselling couples before they start fertility treatment.

Besides the identification of couples with a good prognosis of a spontaneous pregnancy, identification of infertile couples with a very small or no chance of a treatment (in) dependent pregnancy is desirable. Can we draw a line and advise couples to quit or even not start treatment based on ovarian reserve tests using basal Follicle Stimulating Hormone (FSH), Anti-Müllerian Hormone (AMH) and/ or Antral Follicle Count? In the studies we performed, no test on ovarian reserve was accurate enough to be used in IVF-treatment, and treatment of all couples without ovarian reserve testing was found to

generate less distress in couples than testing for ovarian reserve (chapter 6 & 7). Nowadays the Dutch guideline advises that ovarian reserve tests should not be used to predict the chance of a pregnancy, but can be useful to differentiate between women with a low ovarian response and high ovarian response during IVF-treatment and gives information to adjust the dose of medication used for ovarian stimulation (NVOG guideline, 2010). Future research to assess the cost-effectiveness and efficacy (i.e. full-term singleton live birth) of routine use of ovarian reserve test and subsequent use of individualised FSH dosages in predicted poor and high responders as compared to a policy without ovarian reserve test using standard dosages of FSH should be performed. Maybe with the results of these studies we will be able to use ovarian reserve tests as a tool for fine-tuning IVF-treatment and to select couples that will not benefit from IVF treatment.

Samenvatting

Subfertiliteit is gedefinieerd als het uitblijven van een zwangerschap na tenminste 12 maanden onbeschermd vrijen. Ongeveer 9 tot 15% van de paren met kinderwens krijgt volgens deze definitie met subfertiliteit te maken (Beurskens *et al.*, 1995; Boivin *et al.*, 2007). De incidentie van subfertiliteit stijgt in de Westerse Wereld, dit komt met name door het uitstellen van een zwangerschap. Een paar met onvervulde kinderwens zal zich eerst bij de huisarts melden en, als de huisarts beoordeelt dat er inderdaad sprake is van subfertiliteit, verwezen worden naar een gynaecoloog voor het oriënterend fertiliteitsonderzoek ofwel vruchtbaarheidsonderzoek (OFO). Tijdens het OFO worden een aantal standaard onderzoeken verricht: analyse van de kwaliteit van het zaad, testen van de doorgankelijkheid van de eileiders en beoordelen of er een eisprong plaats vindt. Het onderzoek heeft een tweeledig doel, ten eerste het vinden van een verklaring waarom een zwangerschap uitblijft, zodat beoordeeld kan worden of deze stoornis opgeheven of gecorrigeerd kan worden. Ten tweede kan de kans op een spontane zwangerschap in het komende jaar berekend worden aan de hand van de resultaten van het onderzoek met behulp van een prognostisch rekenmodel (NVOG-richtlijn, 2004).

De verklaring voor het uitblijven van een zwangerschap is bij 20% van de paren een probleem met de eisprong, bij 25% een slechte/ matige zaadkwaliteit (= mannelijke subfertiliteit), bij 25% een probleem met de eileiders en bij 30% wordt geen verklaring voor het uitblijven van de zwangerschap gevonden (= onverklaarde subfertiliteit) (Collins and Van Steirteghem, 2004). Als er een verklaring gevonden wordt voor het niet zwanger worden, dan wordt een behandeling voorgesteld waarmee de kans op een zwangerschap vergroot wordt. Bij afgesloten eileiders bijvoorbeeld wordt voorgesteld om een in vitro fertilisatie behandeling (IVF) ofwel reageerbuis behandeling te starten. Indien er sprake is van onverklaarde subfertiliteit of milde mannelijke subfertiliteit (ongeveer 50% van de paren), dan zijn afwachtend beleid (geen behandeling), intra-uteriene inseminatie (IUI) met of zonder stimulatie van de eierstokken of IVF de mogelijke opties. Behandeling met IUI of IVF leidt niet altijd tot een grotere kans op zwangerschap of levendgeborene dan een afwachtend beleid. IUI en IVF zijn wel duurder dan afwachtend beleid en hebben een groter risico op ongewenste uitkomsten, zoals meerlingzwangerschappen of overstimulatie van de eierstokken (Bensdorp *et al.* 2007; Pandian *et al.* 2012; Veltman-Verhulst *et al.* 2012).

Om subfertiele paren goed te kunnen counselen, is het noodzakelijk om te weten met welke ideeën en met welk doel ze naar een dokter gaan. Het is daarom noodzakelijk te weten wat de succesansen van de vruchtbaarheidsbehandelingen zijn en hoeveel paren stoppen met de behandeling in de loop van de tijd. In dit proefschrift zijn verschillende onderdelen van het traject wat een paar met onvervulde kinderwens kan doorlopen nader bekeken.

Hoofdstuk 1 is de inleiding en bevat achtergrond informatie over dit proefschrift.

Het voortplantingstraject wat een paar met kinderwens aflegt is te vergelijken met een reis zonder routebeschrijving: op een gegeven moment kan een paar besluiten aan deze reis te beginnen, maar ze weten dan niet over welke weg ze gaan reizen, hoe lang de reis gaat duren en of de gewenste bestemming, zwanger worden en een gezond kind krijgen, wel gehaald gaat worden. Een substantieel aantal paren onderbreekt deze reis tijdelijk of geheel voordat de gewenste bestemming bereikt is. In dit proefschrift zijn verschillende stations van de reis die paren met kinderwens kunnen afleggen onder de loep genomen.

In **hoofdstuk 2** worden de resultaten beschreven van een interview met 38 subfertiele paren (38 vrouwen en 28 mannelijke partners) die verwezen werden naar het Máxima Medisch Centrum voor het vruchtbaarheidsonderzoek. Deelnemers aan het interview vonden het belangrijker om een verklaring voor het uitblijven van een zwangerschap te vinden tijdens het OFO, dan om zwanger te worden. Gezamenlijke besluitvorming had de voorkeur van 95% van de deelnemers. De meerderheid van de deelnemers wilde zoveel mogelijk weten van het vruchtbaarheidsonderzoek (73%), vruchtbaarheidsbehandeling (70%) en verminderde vruchtbaarheid (62%). Vrouwen rapporteerden dat ze meer weten van de vruchtbaarheidsonderzoeken dan mannen. Aan de deelnemers werd een keuzemodel voorgelegd, waarin stapsgewijs de kans op een complicatie en de kans op het vinden van een verklaring tijdens het vruchtbaarheidsonderzoek werden veranderd. De voorkeur van de helft van deelnemers veranderde van 'beginnen met het onderzoek' naar '6 maanden langer afwachten' bij > 10% kans op een complicatie, en bij < 24% kans op het vinden van een verklaring. De informatie verkregen uit deze interviews kan gebruikt worden voor het counselen van subfertiele paren die verwezen worden voor het vruchtbaarheidsonderzoek.

In **hoofdstuk 3** is gekeken naar uitval van paren op de IVF wachtlijst van het academisch ziekenhuis van Maastricht. Van 674 paren, die achtereenvolgens op de wachtlijst zijn geplaatst in de periode van juni 2000 tot juli 2003, is de omvang van de uitval bekeken

en de reden voor deze uitval achterhaald. De informatie werd verzameld in 2005 en 2008. Van de 674 paren op de wachtlijst is 87% begonnen met de IVF behandeling en 13% van de paren niet. Van de 86 paren die niet gestart zijn met de behandeling, werd van 85 follow-up informatie verkregen (98,8%). Redenen voor de uitval waren de volgende: ontstaan spontane zwangerschap bij 32 paren (37%), persoonlijke reden om te stoppen bij 31 paren (36%) en medisch advies om niet te starten met behandeling bij 23 paren (27%). Het merendeel van de spontane zwangerschappen trad in de eerste 3 maanden na het plaatsen op de wachtlijst op (30 van de 32, 94%). Van de overige 54 paren zijn er nog 4 zwanger geworden bleek uit de follow-up.

Uit deze gegevens hebben we geconcludeerd dat 13% van de paren uitvalt op een IVF-wachtlijst van 6 maanden en dat het optreden van een zwangerschap de belangrijkste reden voor deze uitval is. Het merendeel van deze zwangerschappen trad op binnen 3 maanden na het plaatsen op de wachtlijst, dit suggereert dat psychologische factoren als afname van stress hier mogelijk een rol bij spelen.

In **hoofdstuk 4** zijn gegevens van de 588 paren, die tussen juni 2000 en juli 2003 in het academisch ziekenhuis van Maastricht met de eerste IVF behandeling zijn gestart, geanalyseerd. Zowel van de paren die het gehele behandeltraject hebben doorlopen als van de paren die met de behandeling zijn gestopt, zijn gegevens verzameld om het effect van stoppen met de behandeling op de gerapporteerde zwangerschapskans van de behandeling te kunnen berekenen. De gegevens van de 588 paren die gestart zijn met de behandeling werden verzameld (verder geduid als de 'behandelde groep'). Door middel van geretourneerde schriftelijke vragenlijsten, zo nodig aangevuld met informatie van de huisarts, werd van 91% van de paren die gestopt zijn met de behandeling een complete follow-up verkregen. De gegevens van de 'behandelde groep' aangevuld met deze follow-up gegevens wordt beschouwd als de 'complete groep'.

Met betrekking tot de prognose van paren die stoppen met de behandeling worden de volgende aannames gedaan om de kans op zwangerschap van de behandeling in opeenvolgende cycli te kunnen berekenen: (I) paren die stoppen hebben geen kans op een zwangerschap, (II) paren die stoppen hebben dezelfde kans op een zwangerschap als paren die doorgaan met de behandeling en (III) paren die stoppen met behandeling op advies van de arts (medische reden) hebben geen kans op een zwangerschap en paren die om een andere reden stoppen hebben dezelfde kans op een zwangerschap als degene die doorgaan met de behandeling. De cumulatieve zwangerschapskans kan op basis van opeenvolgende behandelcycli of op basis van de verstreken tijd na het starten van de behandeling berekend worden. De cumulatieve zwangerschapskansen voor de

‘behandelde groep’ zijn berekend voor 3 opeenvolgende cycli en voor de behandelduur, allen met gebruik van de 3 genoemde aannames. De cumulatieve zwangerschapskans voor 3 opeenvolgende cycli en behandelduur zijn ook berekend voor de ‘complete groep’. De berekende cumulatieve zwangerschapskansen van de ‘behandelde groep’ hebben we vergeleken met die van de ‘complete groep’.

De cumulatieve zwangerschapskans na 3 opeenvolgende cycli varieerde van 63-71% in de ‘behandelde groep’ en was 65% in de ‘complete groep’. De cumulatieve zwangerschapskans 9 maanden na starten van de behandeling varieerde van 54-59% in de ‘behandelde groep’ en was 55% in de ‘complete groep’. De zwangerschapskans van de paren die gestopt waren met de behandeling was 14% (95% betrouwbaarheidsinterval 8, 22%).

We hebben geconcludeerd dat de cumulatieve zwangerschapskans voor een IVF behandeling, waar geen follow-up van uitgevallen paren bekend is, berekend dient te worden met de aanname dat uitgevallen paren een zwangerschapskans hebben die ligt tussen geen kans op zwangerschap en dezelfde kans als paren die doorgaan met behandeling. Verder liet deze studie zien dat paren die stoppen met IVF behandeling 14% kans op een zwangerschap hebben.

In **hoofdstuk 5** is het effect van paren die stoppen met vruchtbaarheidsbehandeling (IUI/ IVF/ ICSI) op de gerapporteerde zwangerschapskans in de volgende cyclus geëvalueerd. In wetenschappelijke artikelen over vruchtbaarheidsbehandelingen is mogelijk een te optimistische weergave van de resultaten van deze behandeling als er een veel paren uitvallen tijdens de behandeling voordat ze zwanger zijn geworden. Er zijn geen studies gedaan die het effect van uitval op het zwangerschapscijfer onderzocht hebben. Wij hebben in medische zoekmachines gezocht naar onderzoeken die gegevens gepubliceerd hebben over het resultaat van opeenvolgende behandelcycli van zowel IUI als IVF/ICSI behandeling. Onderzoeken waren geschikt voor de analyse als het mogelijk was om zwangerschaps- en uitvalpercentages per cyclus te kunnen berekenen. Voor elke behandelcyclus is het uitvalspercentage en zwangerschapspercentage berekend. De analyse werd apart gedaan voor IUI en voor IVF/ ICSI behandeling. Met behulp van een rekenkundige methode (de Pearson correlatie coëfficiënt) werd berekend of het uitvalspercentage en zwangerschapspercentage met elkaar samenhangen. In totaal hebben we gegevens geanalyseerd van 17.555 paren die samen 52.003 IUI behandelcycli hebben ondergaan en van 45.934 paren die 99.345 IVF/ICSI behandelcycli hebben ondergaan. De zwangerschapspercentages voor IUI behandelingen varieerden van 0% tot 50% per cyclus (gemiddeld 8%; 95% betrouwbaarheidsinterval (BI) 7,9) en de

uitvalpercentages varieerden 4% tot 85% per cyclus (gemiddeld 30%; 95% BI 27,33). Voor IVF/ICSI varieerden de zwangerschapspercentages van 7% tot 36% per cyclus (gemiddeld 18%; 95% BI 16, 19) en de uitvalpercentages varieerden van 7% tot 69% per cyclus (gemiddeld 36%; 95% BI 32, 41).

We zagen een significante daling van de zwangerschapspercentages in opeenvolgende IUI en IVF/ ICSI behandelcycli. Er bleek geen statistisch significant effect te zijn van het uitvalpercentage in de ene cyclus op het zwangerschapspercentage in de daarop volgende behandelcyclus. Dat houdt in dat een hoger uitvalpercentage in een behandelcyclus niet leidt tot een hoger zwangerschapspercentage in de daarop volgende behandelcyclus. Dit werd zowel voor de IUI als IVF behandeling gevonden. We concludeerden dan ook dat overschatting van zwangerschapspercentages bij IUI en IVF/ ICSI behandeling mee lijkt te vallen.

In **hoofdstuk 6** is gekeken of het testen van de capaciteit van de eierstokken ('ovariële reserve') voor het starten van een eerste IVF behandeling een toegevoegde waarde heeft. Er werd een beslissingsmodel ontwikkeld voor paren die in aanmerking komen voor IVF behandeling. Drie mogelijke scenario's werden geëvalueerd: (I) ontzeg IVF zonder ovariële reserve testen, (II) test ovariële reserve en start IVF als ovariële reserve toereikend is en (III) start IVF zonder ovariële reserve testen vooraf en ontzeg verdere IVF behandeling indien er een slecht behandelresultaat is in die 1e cyclus. De uitkomstmaat van de behandeling was de geboorte van een kind. De waardering van de combinatie van de gekozen strategie en de verkregen uitkomst werd uitgedrukt op een stress-schaal in eenheden 'vergeefs uitgevoerde IVF cycli'. Uitgangspunten waren dat terecht starten IVF behandeling en terecht ontzeggen IVF behandeling geen stress oplevert. De stress die ontstaat in het geval dat IVF behandeling niet gestart is, terwijl een zwangerschap ontstaan zou zijn, is vervolgens gespecificeerd als de verhouding tussen de verwachte stress bij onterecht ontzeggen IVF en de verwachte stress bij onterecht starten IVF ('stress ratio'). Om realistische schattingen te kunnen maken voor deze 'stress ratio's' hebben zowel patiënten als dokters gestructureerde vragenlijsten ingevuld. De waarde die gehecht wordt aan ovariële reserve testen hangt sterk af van de verwachte zwangerschapskans na IVF en van de waardering van onjuiste beslissingen aan de hand van de testuitslag. Uit scenario's met realistische zwangerschapskansen na IVF behandeling met de bepaalde 'stress ratio's' bleek dat behandelen van alle paren zonder ovariële reserve testen te verrichten minder stress opleverde dan het testen van de ovariële reserve. De testeigenschappen van de methoden die gebruikt worden om ovariële reserve te testen zouden eerst moeten verbeteren voordat het testen een

toegevoegde waarde heeft. Gebaseerd op de beslisanalyse, waarbij de huidige test eigenschappen en de preferenties van patiënten en dokters zijn meegenomen, lijkt testen van ovariële reserves nog niet bruikbaar in de praktijk.

In **hoofdstuk 7** zijn de resultaten gepresenteerd van een meta-analyse waarin de nauwkeurigheid van multivariabele modellen, die de ovariële reserve testen en kans op zwangerschap voorspellen bij IVF behandeling, vergeleken wordt met de antrale follikeltelling als voorspellende test. In de medische online zoekmachines zijn artikelen gezocht waarin resultaten zijn beschreven van multivariabele modellen om ovariële reserve te testen bij vrouwen die een IVF behandeling ondergaan. Geschikte onderzoeken werden geselecteerd, de sensitiviteit (= kans op een terecht positieve testuitslag) en specificiteit (= kans op een terecht negatieve testuitslag) van de testen werden berekend voor het voorspellen van ovariële reserve en voor het voorspellen van uitblijven van een zwangerschap. De mate waarin de test de uitkomst goed voorspelt, werd weergegeven in een 'Receiver Operating Characteristic (ROC)' grafiek. Deze ROC grafiek werd vergeleken met de ROC grafiek van de antrale follikeltelling, die beschouwd werd als de beste voorspeller van ovariële reserve op het moment. Er werden 11 onderzoeken gevonden met gegevens over multivariabele modellen om ovariële reserves te testen. Alle onderzoeken rapporteerden over de voorspelling van slechte ovariële reactie op stimulatie, maar geen enkele over het optreden van zwangerschappen. De sensitiviteit voor het voorspellen van een slechte ovariële reactie op stimulatie varieerde van 39-97% en de specificiteit van 50-96%. Logistische regressie-analyse liet zien dat cohort studies een beter discriminerend vermogen hebben dan case-control studies. Derhalve werd de verdere analyse verricht met de gegevens van de cohort studies. Voor deze cohort studies werd een samenvattende ROC grafiek gemaakt en deze had dezelfde vorm als de ROC grafiek die in een ander onderzoek voor de antrale follikeltelling was gemaakt. Hieruit concludeerden wij dat de nauwkeurigheid waarmee multivariabele modellen de ovariële reactie voorspellen bij vrouwen die een IVF behandeling ondergaan vergelijkbaar is met de nauwkeurigheid van antrale follikeltelling alleen. Er waren geen gegevens beschikbaar over de capaciteit van deze modellen om het ontstaan van een zwangerschap of het geboren worden van een levend kind te kunnen voorspellen. Op basis van de gegevens uit deze meta-analyse is er geen bewijs dat het gebruik van meer dan 1 onderzoek voor het testen van ovariële reserve nuttig is.

Appendices

Dankwoord
Curriculum Vitae

Dankwoord

“Starting something is the best way to finish something”

Toen ik de eerste stappen op wetenschappelijk gebied zette, had ik nog niet het idee dat die zouden leiden tot dit boekje. Ik ben vaak weer opnieuw begonnen, maar het is nu echt af! Veel mensen hebben mij geholpen om dit resultaat te bereiken.

Mijn dank gaat in het bijzonder uit naar de paren met kinderwens, al dan niet bezig met een behandeling, die deelgenomen hebben aan de onderzoeken.

Mijn promotoren Prof. Dr. J.L.H. Evers en Prof. Dr. B.W.J. Mol

Hans, ik zie me nog zitten bij ons kennismakingsgesprek aan het begin van de opleiding. Korting kon ik niet meer regelen, maar ik dacht wel wat wetenschap te gaan doen in de opleiding. Dat werd meer naast dan in de opleiding, maar het resultaat is er! Vooral tijdens de laatste loodjes gaf je me precies de juiste zet in de rug waardoor ik me voor(t) kon blijven bewegen ofwel pro-movere. Je reacties kwamen meestal snel en waren scherp. Dank daarvoor!

Ben Willem, zonder jouw inzet was dit boekje er niet geweest. Ik herinner me nog de eerste afspraak met je in het MMC Veldhoven. Ik was in net Maastricht met de opleiding begonnen en kreeg het advies om met jou te gaan praten, als ik op wetenschappelijk gebied wat wilde gaan doen. Zo gezegd zo gedaan. Vele afspraken volgden, op de meest onmogelijke tijdstippen, live, per mail, telefonisch of via skype zonder beeld. Je vergeleek me eens met een wals: traag maar wel vooruit bewegend. Dan zal het nu ook wel even duren voordat ik tot stilstand kom. Enorm veel dank voor je vertrouwen en uithoudingsvermogen! Geniet van je nieuwe uitdaging in Adelaide.

Leden van de beoordelingscommissie, **professor de Die-Smulders, professor Braat, professor Dirksen, professor Helmerhorst en professor Schouten**, ik dank u allen voor het beoordelen van mijn proefschrift.

Verpleegkundigen van de IVF afdeling in het azM bedankt voor de hulp om de follow-up van de IVF dropouts nagenoeg compleet te krijgen. Fertiliteit-assistenten uit het MMC, **Marja en José**, bedankt voor het benaderen van de paren met kinderwens voor de

interviews. **Mede-auteurs**, zonder jullie hulp was ik niet zover gekomen. **Jolande Land, Angelique van Dongen, Femke Mooren-Luttjeboer, Brent Opmeer, Sjors Coppus en Eline Dancet** bedankt voor jullie inzet op diverse onderdelen bij de verschillende stukken.

Gynaecologen van het **academisch ziekenhuis Maastricht** (inmiddels MUMC+) en gynaecologen van het **Máxima Medisch Centrum**. Dankzij alles wat ik in mijn opleiding van jullie geleerd en afgeleerd heb, ben ik de gynaecoloog geworden die ik nu ben en daar ben ik dankbaar voor. Ook dank aan de **mede-assistenten** voor de gezellige jaren die we samen tijdens de opleiding hebben doorgebracht in Maastricht en Veldhoven/Eindhoven.

Maatschap gynaecologie Slingeland Ziekenhuis, **Kirsten, Eveline, Frans, Walter-Jacques en Pim**. Wat was het fijn om in november 2010 als chef de clinique bij jullie te komen werken. En wat voelt het goed om me nu jullie maat te kunnen noemen. Ik waardeer jullie aanwezigheid op 27 maart enorm.

Verpleegkundigen, verloskundigen (in 1e en 2e lijn werkzaam), **medewerksters** op de **polikliniek** en op **afdeling B2** van het Slingeland Ziekenhuis. Fijn dat we nog langer samen gaan werken!

Mijn paranimfen, Sascha en Minouche. Lieve **Sascha**, vanaf de 1e dag dat jij naar het van Maerlant kwam, was er een klik door jouw humor en ongedwongen gezelligheid. Ondanks mijn vele verhuizingen is die klik gebleven. Op andere bijzondere momenten heb ik al eens naast jou mogen staan. Super dat je 27 maart naast mij staat. Lieve **Minouche**, bij bijna alle life-events nauw betrokken, en die ene keer had je wel een heel goed excuus om er niet bij te zijn. Jouw stelligheid maakt me wel eens jaloers. Achter die stelligheid heb ik een hele warme, oprechte en bijzondere vrouw leren kennen. Lief en leed hebben we gedeeld de laatste jaren. Ik reken erop dat we dat nog vele jaren blijven doen.

Mijn lieve vriendinnen **Brigit, Anke, Jolanda, Bernadet, Leonie, Baukje, Maaïke en Ingrid**. Vriendschappen ontstaan in verschillende fasen van mijn leven, maar in staat gebleken te blijven bestaan. Ik heb de illusie dat ik binnenkort meer tijd heb om jullie allemaal weer vaker te gaan zien nu dit project ten einde is. Laten we snel wat afspreken!

Joop en Gerry, onze lieve trouwe oppassen. Fijn dat jullie er altijd zijn als wij weer met het werk bezig zijn, ook al komt het verzoek weer eens op het laatste moment.

Mijn schoonfamilie **Milly, Swendley, Sandra, Shanice** en **Shandro, Garly, Natascha, Juliëne, Sylvana** en **Shania**. We zien elkaar niet vaak, maar als we elkaar zien is het gezellig.

Lieve zussen **Ruth** en **Sara**. Wat is het een voorrecht om twee zulke geweldige zussen te hebben. Zomaar even bij elkaar binnen vallen, is er niet meer bij door onze woonplaatsen. Maar nu hebben we wel echt de tijd voor elkaar als we elkaar zien. De zussendagen/-weekenden houden we in ere!

Ruth & Casper met **Vera, Suus** en **Noor** en **Sara & Marcel** met **Floor** en **Maud**, fijn dat jullie deur altijd voor ons open staat. We blijven graag komen.

Casper, VormMeester van de cover en layout van dit proefschrift. Geweldig werk heb je geleverd!

Lieve **Lia** en **Dré**, jullie onvoorwaardelijke steun en grenzeloze vertrouwen in mij hebben er zeker aan bij gedragen dat ook dit project tot een goed einde is gekomen. Ik heb het enorm getroffen met mijn ouders! Lieve **Lia**, als ik weer eens een schrijf-vakantie-week had, was de aanwezigheid van oma de reden dat de kinderen het prima vonden dat mama boven zat. Heerlijk om alles dan zo uit handen te kunnen geven. Lieve **Dré**, op het moment dat ik er echt mee wilde stoppen heb jij me opnieuw moed ingeprepraat. En zie: het heeft geholpen!

Lieve **Eline** en **Fabian**, fijn hè “het boekje” is nu echt af! Nu hoef ik niet meer boven te werken. Jullie afleiding tijdens het schrijven was vaak een welkom moment om even te kunnen ontspannen. Het is een genot om jullie op te zien groeien als vrolijke, sociale, ondeugende, eigenwijze en slimme kinderen. Het is super om jullie mama te zijn!

Mi Amor **Ashley**, ontzettend veel dank voor de ruimte die je me gegeven hebt om, in onze spaarzame gezamenlijke vrije tijd, aan dit proefschrift te kunnen werken. Jouw rust en geduld zijn meer dan eens goed van pas gekomen. Het volgende project wordt een gezamenlijk project!

Mi stimabo hopi X

Curriculum Vitae

Tamara Verhagen werd op 4 februari 1975 geboren in Weert en groeide op in Veldhoven. In 1993 behaalde ze haar Gymnasium diploma aan het Van Maerlantlyceum te Eindhoven. Na 2 jaar gezondheidswetenschappen gestudeerd te hebben aan de Radboud Universiteit Nijmegen werd ze in 1995 ingeloot voor de studie geneeskunde in Nijmegen. Datzelfde jaar behaalde ze haar propedeuse en in november 2000 studeerde ze cum laude af als basisarts om vervolgens als ANIOS aan de slag te gaan bij de vakgroep obstetrie en gynaecologie van de Gelderse Vallei te Ede. In 2002 ging ze als ANIOS werken in het Radboudumc te Nijmegen. In maart 2003 startte ze met de opleiding tot gynaecoloog in het Maastricht Universitair Medisch Centrum (opleiders prof. dr. J.L.H. Evers en prof. dr. G.G. Essed) en woonde ze een aantal jaar met veel plezier in het zuiden van het land. In Maastricht werd tevens de basis voor dit proefschrift gelegd. In oktober 2005 verhuisde ze terug naar Eindhoven om haar opleiding te vervolgen in het Máxima Medisch Centrum Veldhoven/ Eindhoven (opleiders prof. dr. S.G. Oei en dr. M.Y. Bongers). Op 17 januari 2010 was ze klaar met de opleiding en is, na een korte periode gewerkt te hebben in het Elkerliek Ziekenhuis te Helmond, als chef de clinique naar het Slingeland Ziekenhuis in Doetinchem vertrokken. Per 1 november 2013 is daar ze aangesteld als maatschapslid. Tamara is getrouwd met Ashley Meri n en samen hebben ze 2 prachtige kinderen: Eline (geboren in 2006) en Fabian (geboren in 2008).



Subfertility
decision-to-decision path

 *point of treatment*

point of decision

E.M. Verhagen

