Prediction Models in Reproductive Medicine

Jan Willem van der Steeg
2008

The studies presented in this thesis were supported by funds from the Netherlands Organisation for Health Research and Development (ZonMw, project number 945/12/002).

Financial support for printing of the thesis was provided by:

Dept. of Public Health, Erasmus MC
Vrouwenkliniek AMC
Stichting Gynaecologische Endocrinologie en Kunstmatige Humane Voortplanting
Freya, vereniging voor mensen met vruchtbaarheidsproblemen

Abbott BV	Medical Dynamics	Organon
Bayer Schering Pharma	Mediparc BV	Sigma Medical BV
Cytyc	Memidis Pharma BV	Will Pharma
Esaote	Merck Serono	Solvay Pharmaceuticals
Ferring BV	Nycomed BV	
Goodlife Fertility	Olympus	

Prediction models in Reproductive Medicine

Thesis, Erasmus MC, University Medical Center Rotterdam – with references – with summary in Dutch

Printed by: Uitgeverij Ponsen en Looijen BV
Cover design by: Mirelle van Beers
Lay-out: Redactie- en projectbureau Können

ISBN/EAN: 978-90-6464-239-5

Prediction Models in Reproductive Medicine

Prognostische modellen in de voortplantingsgeneeskunde

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 16 mei om 11.00 uur

door

Jan Willem van der Steeg

geboren te Eindhoven

Promotiecommissie

Promotoren:	Prof.dr. J.D.F. Habbema Prof.dr. F. van der Veen
Overige leden:	Prof.dr. E.A.P. Steegers Prof.dr. G.J. Bonsel Prof.dr. B.C.J.M. Fauser
Copromotoren:	Prof.dr. B.W.J. Mol Dr. M.J.C. Eijkemans

Aan mijn ouders en Marieke

Contents

chapter (01)

GENERAL INTRODUCTION

General introduction

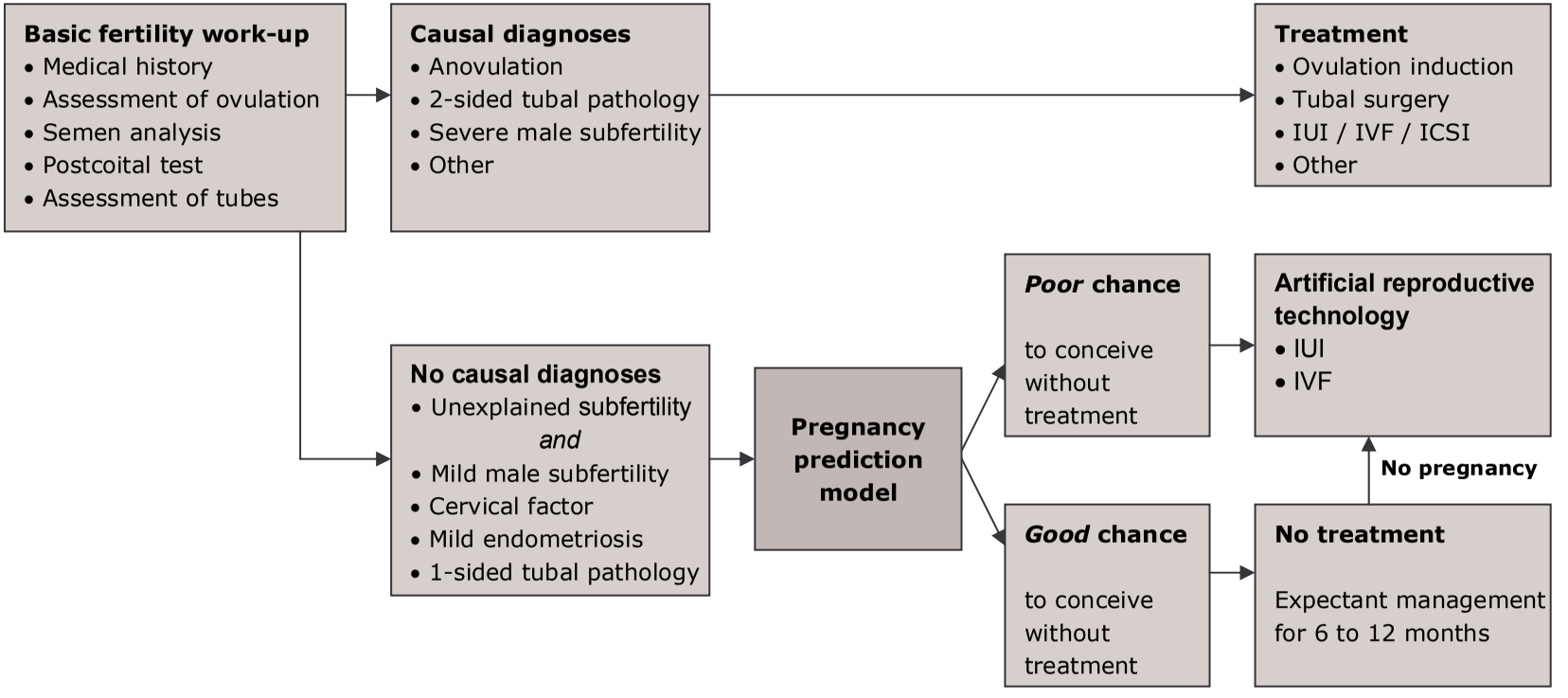
Subfertility is defined as failure to conceive after at least one year of unprotected intercourse and occurs in approximately one out of ten couples (Gibbons 1911; Hull *et al.,* 1985; Wang *et al.,* 2003; Gnoth *et al.,* 2003). Annually, almost 20 000 couples in The Netherlands are confronted with subfertility.

The majority of these subfertile couples seek medical help (Evers 2002). Current fertility work-up consists of several tests and starts with a detailed medical history, followed by assessment of ovulation, semen analysis, postcoital test (PCT), and assessment of the Fallopian tubes (National institute for clinical excellence 2004; Dutch Society of Obstetrics and Gynaecology 2004). After the fertility work-up, diagnoses such as anovulation, tubal pathology, endometriosis, cervical factor, male subfertility, or sexual dysfunction can be made. In 20% of couples however, no cause for the subfertility is found (Hull *et al.,* 1985; Evers 2002; Collins and Van Steirteghem 2004).

In couples with a diagnosis severely impairing the chance to conceive, a causal treatment will be offered. For example, subfertility due to anovulation will be treated initially with ovulation induction. Couples of whom the women are diagnosed with double-sided tubal pathology will be treated with tubal surgery or *in vitro* fertilization (IVF), depending on the severity of tubal damage. In couples in whom the men is diagnosed with severely impaired semen parameters, the treatment of first choice will consist of IVF with or without intracytoplasmic sperm injection (ICSI).

In couples in whom no diagnosis could be made, i.e. couples with unexplained infertility, or couples in whom the man is diagnosed with mildly impaired semen quality, or in whom the woman is diagnosed with a cervical factor, mild endometriosis or one-sided tubal pathology, the prognosis for spontaneous pregnancy will determine whether treatment is indicated, because interventions should only be used when their success rates clearly exceed the probability of a spontaneous pregnancy (Te Velde and Cohlen 1999; Wasson *et al.,* 1985). It is therefore very important to develop and use pregnancy prediction models in order to distinguish couples with a poor probability of spontaneous pregnancy, in whom IUI or IVF is mandatory, from couples who still have a good probability to conceive spontaneously (figure).

Figure Diagnostic and prognostic flow chart for subfertile couples.



Such prediction models for pregnancy may be a potential tool to prevent unnecessary treatment. In the past decade, several prediction models to calculate the probability of a spontaneous pregnancy were developed. The three most relevant prediction models are the models of Eimers, Collins and Snick (Eimers *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997).

The first model was developed by Eimers *et al*. at the University Medical Hospital Utrecht in 1994. The model was based on a mixed population of a secondary and third care patient group. Over 900 couples presenting with subfertility in the period 1974 and 1984 were evaluated retrospectively. The second model was developed by Collins *et al*., and was based on more than 2000 subfertile couples in third care fertility centres in Canada between 1984 and 1987. The third model was developed by Snick *et al*. and was based on more than 700 subfertile couples that visited a secondary care fertility centre in Walcheren, The Netherlands, between 1985 and 1993. All models included data collected during the basic fertility work-up.

In 2004, Hunault *et al.* combined the models of Eimers, Collins and Snick into a new model for spontaneous pregnancy and cross-validated the results on the three original datasets (Hunault *et al.,* 2004). This new model is based on the variables female age, duration of subfertility, subfertility being primary or secondary, referral status (being referred by the general practitioner or another gynaecologist), percentage motile sperm and result of the postcoital test. This synthesis model may predict spontaneous pregnancy more accurately, as it is based on the combined data of the older models.

Aim of the thesis

At the time the studies described in this thesis were designed, the use of prediction models for pregnancy was not incorporated in fertility care, and fertility guidelines did not even mention the use of prediction models (Boston (MA) 2003; National institute for clinical excellence 2004). We were interested in the need for prediction models in reproductive medicine. If gynaecologists were able to predict probabilities of pregnancy reliably, based on their clinical experience and gut feeling, prediction models would not be necessary. In contrast, if gynaecologists were not able to estimate pregnancy chances accurately and reliably, there might be a need for prediction models. To explore this issue, we designed a survey among gynaecologists in which they had to estimate pregnancy chances in couples that varied in prognostic profile.

Secondly, we were interested in the performance of prediction models in a general subfertile population. Therefore, we designed a multicentre prospective cohort study among subfertile couples that had completed their basic fertility work-up. Subsequently, a probability of spontaneous ongoing pregnancy was calculated, which determined whether the couple should be counselled for expectant management or for treatment. These probabilities were evaluated after follow-up was completed in terms of discrimination and accuracy.

The generated dataset was not only used to validate current prediction models, but also to evaluate whether factors that at present are not incorporated in the models, could improve the performance of the prediction model. Factors considered were body mass index, follicle stimulating hormone, a detailed analysis of reproductive history, and semen analysis. Finally, a new prediction model for spontaneous pregnancy was developed.

The aims of this thesis can be formulated in the following six questions:

1. What is the right moment to perform the basic fertility work-up with respect to duration of child wish?

2. Are gynaecologists able to make reliable and reproducible predictions of pregnancy chances for spontaneous pregnancy and pregnancy after IUI and IVF?

3. Which factors of the basic fertility work-up play currently an important role in clinical decision-making by gynaecologists?

4. How does the prediction model for spontaneous pregnancy of Hunault perform in a general subfertile population?

5. Can the existing factors of the prediction model for spontaneous pregnancy be refined, with special emphasis on pregnancy history, semen analysis and the postcoital test?

6. What is the prognostic capacity of the body mass index (BMI) and basal follicle stimulating hormone (FSH), potential prognostic factors that are not incorporated in current prediction models?

Outline of the thesis

In **chapter 2** we discuss the right moment to investigate the fertility potential of couples with active child wish. We assess the consequences of performing the basic fertility work-up already after six months instead of after twelve months of unfulfilled child wish. We focus on the conception chances, the false positive diagnosis rate and the additional effect of fertility treatment in case the fertility work-up will be performed in an earlier stage.

In **chapter 3** we report the results of a survey among gynaecologists in the Netherlands. We evaluate their ability to predict pregnancy chances in fictive patient cases.
Firstly, we evaluate the concordance between these gynaecologists with respect to predictions on spontaneous pregnancy, pregnancy after IUI and pregnancy after IVF.
Secondly, we evaluate the concordance between the gynaecologists with respect to their subsequent treatment decisions.

In **chapter 4** we report which factors of the fertility work-up are of importance for gynaecologists both in their estimates of pregnancy chances of treatment-independent pregnancy, IUI and IVF, and in their subsequent treatment decisions. Results of this study are based on a different part of the same survey as described in chapter 3. We also compare the importance of each of these factors as found in this study with their importance as reported in existing prediction models and guidelines.

In **chapter 5** we validate the prediction model for spontaneous ongoing pregnancy (model of Hunault). The external data set consisted of a prospective cohort subfertile couples, collected in 38 hospitals in the Netherlands in the period between 2002 and 2004.
In total, 3021 subfertile couples were included, the women having an ovulatory cycle and two patent tubes, the men having a total motile sperm count over 3 million. We assess the performance of the prediction model by its calibration and its discriminative capacity.

In **chapter 6** we present the results of a large prospective cohort study, as described in chapter 5, on how the outcome (live birth or not (miscarriage, induced abortion)), the localization (intra uterine or ectopic pregnancy) the origin (current or previous partnership) and the mode of conception (with or

without treatment) of a prior pregnancy in subfertile couples are related to the probability of spontaneous ongoing pregnancy. The study included 4445 couples, the women having an ovulatory cycle and at least one patent tube, the men having a total motile sperm count over 3 million.

Chapter 7 focuses on the predictive value of day-3 FSH levels in subfertile women with a regular cycle and at least one patent tube, the men having a total sperm count over 3 million. The study is part of the large cohort study, as described in chapter 5. In total, 3519 women were included. We also evaluate in how many couples fertility treatment changes from expectant management to ART and vice versa when prognostic information of basal FSH is used in the clinical decision-making.

In **chapter 8** we determine the association between obesity and the probability of spontaneous pregnancy in 3029 subfertile ovulatory women who had at least one patent tube, the men having a total sperm count over 3 million. This study was part of the large prospective cohort study described in chapter 5. In the analysis we adjust for potential related factors female age, duration of subfertility, previous pregnancy, referral status (2^{nd} or 3^{rd} care), semen motility, and current smoking of the female and male partner.

In **chapter 9** we evaluate how WHO criteria for semen perform in a general subfertile population. This study included 3345 men without azoospermia, of whom the woman had an ovulatory cycle and at least one patent tube. This study was a part of the large multicenter cohort study, described in chapter 5. Furthermore, we propose new thresholds for semen as derived from a general subfertile population.

In **chapter 10** we report whether the PCT needs to be performed routinely in the fertility work-up. We assess if the results of the PCT can be predicted from the patients' history and the semen analysis. The analysis was performed on the prospective cohort collected by Snick in the period 1985 to 1993, including 522 subfertile couples. We report whether omission of the PCT in couples in whom the PCT result can be predicted would compromise the capacity to predict spontaneous pregnancy from data obtained at the fertility work-up.

In **chapter 11** we discus the results of this thesis and give our view of the future of prediction models in reproductive medicine.

In **chapter 12** we summarize the obtained results of the studies presented in this thesis and present implications for clinical decision-making in the fertility work up of subfertile couples.

References

Bostton (MA). (2003) Infertility. A guide to evaluation, treatment, and counseling. www.guideline.gov.

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Collins JA and Van Steirteghem A. (2004) Overall prognosis with current treatment of infertility. *Human Reproduction Update* 10,309-316.

Dutch Society of Obstetrics and Gynaecology. (2004) Guideline-Basic fertility work-up. *NVOG-richtlijn nr 1*,1-12.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Evers JL. (2002) Female subfertility. *Lancet* 360,151-159.

Gibbons RA. (1911) A lecture on sterility, its etiology and treatment. *London: J&A Churchill*.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, and Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18,1959-1966.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, and Desai KM. (1985) Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 291,1693-1697.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems. www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Te Velde ER and Cohlen BJ. (1999) The management of infertility. *N Engl J Med* 340,224-226.

Wang X, Chen C, Wang L, Chen D, Guang W, and French J. (2003) Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79,577-584.

Wasson JH, Sox HC, Neff RK, and Goldman L. (1985) Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 313,793-799.

HUMAN REPRODUCTION
2005; 20: 2672 - 74

Jan Willem van der Steeg
Pieternel Steures
Peter G.A. Hompes
Marinus J.C. Eijkemans
Fulco van der Veen
Ben Willem J. Mol

chapter (02)

INVESTIGATION OF THE INFERTILE COUPLE

a basic fertility work-up performed within 12 months of trying to conceive generates costs and complications for no particular benefit

Abstract

The current approach of the basic fertility work-up has been questioned recently in this journal. Based on new data on human fecundity, the authors advocated starting the fertility work-up after just six months of trying to conceive instead of the usual twelve months. In women younger than 39 years and with a regular cycle, there are several arguments why the basic fertility work-up should not be done earlier than after 12 months of child wish.

Firstly, 50% of couples who have tried to conceive for six months without success will conceive in the next coming six months without any treatment.

Secondly, the prevalence of fertility diseases is lower in couples who have been trying to conceive for six months as compared with those who have been trying for twelve months. Performance of a fertility work-up at this stage will lead to an increase in false-positive diagnoses compared with performing them at twelve months of subfertility.

Thirdly, fertility treatment will have fewer additional effects in couples with good spontaneous conception prospects (6-12 months child wish), compared with subfertile couples who have poor prospects. At present, none of the available fertility treatments have success rates comparable with no intervention in these women, and postponement of treatment in such couples will prevent complications as ovarian hyperstimulation syndrome and multiple pregnancies.

We argue that the fertility work-up should not be offered to couples with a duration of child wish of < 12 months, except for women with ovulation disorders and women of 39 years and older.

Introduction

The current fertility work-up, for example as recommended in the National Institute for Clinical Excellence (NICE) guidelines, is performed in couples who have an unfulfilled child wish for at least 12 months (National institute for clinical excellence 2004).

Since recent prospective studies on fecundity have shown that ~ 80% of all couples conceives within six months after discontinuation of contraception, and that only 10% of the initial group will conceive within the next six months, the timing of the fertility work-up has been debated (Gnoth *et al.,* 2003; Wang *et al.,* 2003). Brosens *et al.* advocated that female subfertility should be investigated after six months of child wish (Brosens *et al.,* 2004). One of their arguments was that the current definition of subfertility is based on a probability to conceive of 20% per cycle or ~85% per year. According to the new data, this point is already reached at six months, in contrast to previous, mostly retrospective studies, which reported a cumulative pregnancy rate of ~85% at 12 months (Hull *et al.,* 1985; Snick *et al.,* 1997; Evers 2002; Brosens *et al.,* 2004). Gnoth *et al.* also advocated performing the basic fertility work-up at an earlier stage, stating that after six cycles of unprotected intercourse without conception, every second couple is probably subfertile (Gnoth *et al.,* 2003).

In the following sections, we explore four arguments why the fertility work-up should not be offered to couples with a child wish of < 12 months.

The 'spontaneous pregnancy' argument

In the studies by Gnoth *et al.* and Wang *et al.* 80% of the couples conceived successfully within six months, whereas only 10% of patients conceived in the six months thereafter (Gnoth *et al.,* 2003; Wang *et al.,* 2003). The results of these studies were used as an argument for the early evaluation of the reproductive capacity of couples who had not conceived after six months, since 50% of those that had not conceived after six months would not conceive in the next six months.

Now let us flip the coin, and look at it from the other side. The same data implicate that 50% of couples who had not conceived after six months will conceive in the next six months. Pregnancy rates as high as 50% are rarely reported in reproductive medicine. As a matter of fact, we are not

aware of any treatment that has such pregnancy rates, without generating harm, side effects and costs. Thus, whereas used as an argument for starting a diagnostic work-up, the data of Gnoth *et al.* and Wang *et al.* are a strong motive for reassurance of couples who did not conceive after six months, to aim for natural conception for another six months.

The 'false-positive diagnosis' argument

The recently published NICE guidelines recommend performance of ovulation detection, semen analysis as well as hysterosalpingography (HSG) in couples who do not conceive within 12 months (National institute for clinical excellence 2004). Apparently, these tests are valuable for couples who have tried to conceive for > 12 months.

What will happen if we perform these tests after six months of unfulfilled child wish? As an example, we use the HSG. For HSG, the sensitivity is known to be 65% for a specificity of 83% (Mol *et al.,* 1996). When these tests are used in a population of subfertile couples who have been trying to conceive for at least twelve months, the prevalence of tubal pathology in this group will be ~ 20% (Mol *et al.,* 1997). Table 1 shows the two-by-two table that can be constructed if we perform a test with a sensitivity of 65% and a specificity of 80% in a population of 1000 fictitious subfertile couples, where the prevalence of tubal disease is 20%. Out of 200 patients with tubal pathology, HSG will identify 130 cases, whereas out of 800 couples without tubal pathology, HSG will incorrectly point out 160 as having tubal pathology. This means that if we assume all patients with an abnormal HSG are scheduled for subsequent laparoscopy to assess their tubal status in detail, 130 out of 290 patients that will undergo laparoscopy indeed have tubal pathology (44.8%), whereas 55.2% have not.

Table 1 Test results of HSG when it is performed only in subfertile couples (trying to conceive for at least 12 months).

		Laparoscopy		
		Tubal pathology	**No tubal pathology**	**Total**
HSG	**Tubal pathology**	130	160	290
	No tubal pathology	70	640	710
	Total	200	800	1000

Test properties: sensitivity 65% for a specificity of 80%.
Prevalence of tubal pathology is assumed to be 20%.

Now let us consider what happens when the HSG is applied among couples who have been trying to conceive for six months rather than twelve months. From the studies of Gnoth *et al.* and Wang *et al.,* we know that of those couples who are not pregnant after six months, 50% will conceive and have an ongoing pregnancy within the next six months (Gnoth *et al.,* 2003; Wang *et al.,* 2003). These couples will not have relevant tubal pathology, and as a consequence, the prevalence of tubal pathology in a cohort of couples who have not conceived within 6 months of trying is 10%. If we now perform the HSG in 1000 women after six months of whom 100 have tubal pathology, the number of patients with tubal pathology that have been identified by HSG will be 65. The number of patients in which a false-positive diagnosis will be made will be 180 (table 2). Thus the percentage of patients in which subsequent laparoscopy will show tubal abnormalities will drop from 44.8% (130 out of 290) to 26.5% (65 out of 245) or, the other way around, the percentage of patients in which laparoscopy is performed and no tubal pathology is found increases from 55.2% to 73.5%.

Table 2 Test results of HSG where it is performed both in subfertile couples (trying to conceive for at least 12 months) and in couples who are trying to conceive for 6-12 months.

		Laparoscopy		
		Tubal pathology	**No tubal pathology**	**Total**
HSG	**Tubal pathology**	65	180	245
	No tubal pathology	35	720	755
	Total	100	900	1000

Test properties: sensitivity 65% for a specificity of 80%.
Prevalence of tubal pathology is assumed to be 10%.

The above mechanism in which the number of false- positive diagnoses increases due to the fact that the fertility work-up started after 6 months of child wish instead of after 12 months of child wish may also occur in other diagnostic tests, for example the semen analysis. Ombelet *et al.* demonstrated in a comparison between semen parameters and subsequent pregnancy that the predictive performance of these tests is poor. From the receiver operating characteristic (ROC) curves they provided, it can be derived that for normal count (cut-off value 5.6 $x10^6$), and a sensitivity of 62%, the specificity of semen analysis in the prediction of pregnancy will be ~ 82% (Ombelet *et al.,* 1997). Other semen parameters are reported less specific for a sensitivity of 60%. As a consequence, the example of HSG in the diagnosis of tubal pathology is also true for semen analysis in the diagnosis of male factor subfertility.

The 'no-benefit from treatment' argument

The third point is whether early treatment will improve fertility prospects. Virtually all the available evidence on the effectiveness of treatment is based on studies that included couples who were subfertile according to the World Health Organization guidelines, i.e. 'either a woman's inability to conceive and bear a live child or a man's inability to impregnate a woman over a twelve month period of unprotected, regular and normal sexual intercourse.' (Vayena *et al.,* 2002). We do not know whether this evidence can be transferred to couples who are trying to conceive for six months.

Let us consider the example of a couple with unexplained subfertility. According to a randomized clinical trial of Guzick *et al.*, treatment of a couple with unexplained subfertility with superovulation and intrauterine insemination (IUI) would increase the probability of an ongoing pregnancy with a relative risk (RR) of 4.4 (95% confidence interval (CI) 2.5 – 7.9), i.e. from 2.0% per cycle without treatment to 8.7% per cycle with treatment (Guzick *et al.,* 1999). Now we offer the same treatment to a couple who have tried to conceive for six months. A 50% pregnancy rate per six months implies that the spontaneous pregnancy rate per cycle these couples is 11% ($P = 100\% - (100\%-11\%)^6 = 50\%$). Consequently, a RR of 4.4 would imply that the pregnancy rate per IUI cycle should be 48%, corresponding to a success rate of 98% after six months of IUI treatment, which is virtually impossible.

The 'prevention of complications' argument

Finally, there is the prevention of complications argument. Complications in reproductive medicine are relatively rare but, once present, they can have serious consequences for the individuals who were, apart from their unfulfilled child wish, mostly healthy. An example of such a complication is ovarian hyperstimulation syndrome (OHSS). Many measures have been proposed to prevent OHSS, and recently a recommendation has been made for the prevention of OHSS (Practice Committee of the ASRM 2003). Among the measures proposed was that the ovulation induction regimens should be highly individualized, careful monitoring should be carried out and minimum dose and duration of gonadotropin therapy necessary to achieve the therapeutic goal should be used.

Unfortunately, a very obvious way to prevent OHSS, i.e. postponement of treatment, was not mentioned at all. The same accounts for what might be the most important side effect these days, i.e. multiple pregnancies (Fauser *et al.,* 2005). Although single embryo transfer (SET) with cryopreservation of the remaining embryo's results in a considerable reduction of the multiple pregnancy rate, with an acceptable, although somewhat reduced number of singleton pregnancies, postponement of treatment for six months results in a 50% conception rate of singleton pregnancies.

In conclusion, we feel that offering a fertility work-up to couples who did not achieve conception within six months of child wish would lead to an unacceptable number of false-positive diagnoses, treatment of couples that would have a considerable chance of conception without treatment, and unnecessary side effects and costs. We plead for careful counselling of those couples who desire a child, with an emphasis on spontaneous conception chances and the relative value of treatment, followed by an adequate diagnostic work-up and careful counselling once pregnancy has not occur after one year.

An exception can be made for selected groups of patients, e.g. those couples in whom the female partner does not have a regular menstrual cycle or couples in whom the female partner is over 39 years of age. In the latter group, there is no high quality evidence that the postponement of treatment will harm, but cohort studies suggest a strong decrease of the reproductive capacity in these women with a further increase of age (Schwartz and Mayaux 1982; Noord-Zaadstra *et al.,* 1991).

Those who argue that evaluation and subsequent treatment should be performed earlier are invited to demonstrate the value of such interventions in well designed diagnostic and therapeutic studies.

References

Brosens I, Gordts S, Valkenburg M, Puttemans P, Campo R, and Gordts S. (2004) Investigation of the infertile couple: when is the appropriate time to explore female infertility? *Hum Reprod* 19,1689-1692.

Evers JL. (2002) Female subfertility. *Lancet* 360,151-159.

Fauser BCJM, Devroey P, and Macklon NS. (2005) Mulltiple birth resulting from ovarian stimulation for subfertility treatment. *Lancet* 365,1807-1816.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, and Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18,1959-1966.

Guzick DS, Carson SA, Coutifaris C, Overstreet JW, Factor-Litvak P, Steinkampf MP, Hill JA, Mastroianni L, Buster JE, Nakajima ST *et al*. (1999) Efficacy of superovulation and intrauterine insemination in the treatment of infertility. National Cooperative Reproductive Medicine Network. *N Engl J Med* 340,177-183.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, and Desai KM. (1985) Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 291,1693-1697.

Mol BWJ, Swart P, Bossuyt PMM, Van Beurden M, and Van der Veen F. (1996) Reproducibility of the interpretation of hysterosalpingography in the diagnosis of tubal pathology. *Hum Reprod* 11,1204-1208.

Mol BWJ, Swart P, Bossuyt PMM, and Van der Veen F. (1997) Is hysterosalpingography an important tool in predicting fertility outcome? *Fertil Steril* 67,663-669.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems. www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Noord-Zaadstra BM, Looman CW, Alsbach H, Habbema JDF, Te Velde ER, and Karbaat J. (1991) Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *BMJ* 302,1361-1365.

Ombelet W, Bosmans E, Janssen M, Cox A, Vlasselaer J, Gyselaers W, Vandeput H, Gielen J, Pollet H, Maes M *et al*. (1997) Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing. *Hum Reprod* 12,987-993.

Practice Committee of the ASRM. (2003) Ovarian hyperstimulation syndrome. *Fertil Steril* 80,1309-1314.

Schwartz D and Mayaux MJ. (1982) Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS. *N Engl J Med* 306,404-406.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Vayena E, Rowe PJ, and Griffin PD. Current Practices and Controversies in Assisted Reproduction. Report of a meeting on "Medical, Ethical and Social Aspects of Assisted Reproduction" held at WHO Headquarters in Geneva, Switzerland 17-21 September 2001. 2002.

Wang X, Chen C, Wang L, Chen D, Guang W, and French J. (2003) Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79,577-584.

BRITISH JOURNAL OF OBSTETRICS AND GYNAECOLOGY
2006; 113: 825 - 831

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Patrick M.M. Bossuyt
Peter G.A. Hompes
Fulco van der Veen
Ben Willem J. Mol

chapter (03)

DO CLINICAL PREDICTION MODELS
improve concordance of treatment decisions in reproductive medicine?

Abstract

Objective To assess whether the use of clinical prediction models improves concordance between gynaecologists with respect to treatment decisions in reproductive medicine.
Design We constructed 16 vignettes of subfertile couples by varying fertility history, postcoital test, sperm motility, follicle stimulating hormone level and Chlamydia antibody titre.
Setting Thirty-five gynaecologists estimated three probabilities, i.e. the one-year probability of spontaneous pregnancy, the pregnancy chance after intrauterine insemination (IUI) and the pregnancy chance after *in vitro* fertilisation (IVF). Subsequently they proposed therapeutic regimens for these 16 fictional couples, i.e. expectant management, IUI or IVF.
Three months later, the participant gynaecologists again had to propose therapeutic regimes for the same 16 fictional cases but this time accompanied by pregnancy chances obtained from prediction models: predictions on spontaneous pregnancy, IUI and IVF.
Population Thirty-five gynaecologists working in academic and non-academic hospitals in the Netherlands.
Methods Setting section.
Main outcome measures The concordance between gynaecologists of probability estimates, expressed as interclass correlation coefficient (ICC) and the concordance between gynaecologists of treatment decisions, analysed by calculating Cohen's kappa (κ).
Results The gynaecologists differed widely in estimating pregnancy chances (ICC: 0.34). Furthermore, there was a huge variation in the proposed therapeutic regimens (κ: 0.21). The treatment decisions made by gynaecologists were consistent with the ranking of their probability estimates. When prediction models were used, the concordance (κ) for treatment decisions increased from 0.21 to 0.38. The number of gynaecologists counselling for expectant management increased from 39% to 51%, whereas counselling for IVF dropped from 23% to 14%.
Conclusions Gynaecologists differed widely in their estimation of prognosis in 16 fictional cases of subfertile couples. Their therapeutic regimens showed likewise huge variation. After confrontation with prediction models in the same 16 fictional cases, the proposed therapeutic regimens showed only slightly better concordance. Therefore a simple introduction of validated prediction models is insufficient to introduce concordant management between doctors.

Introduction

The challenge of modern subfertility treatment is to offer a tailored treatment to individual subfertile couples. Both early treatment in couples with high chances of spontaneous pregnancy and unnecessary delay of treatment (expectant management) in couples with poor chances of spontaneous pregnancy should be avoided, resulting in cost-effectiveness, reduced multiple pregnancies and reduced complications after assisted reproduction technique (ART) (Gnoth *et al.,* 2005).

Can gynaecologists predict pregnancy chances based on their clinical experience and gut feeling? One study reported a good concordance between gynaecologists in predicting spontaneous pregnancy chance but a poor concordance in predicting pregnancy after *in vitro* fertilisation (IVF) (Wiegerinck *et al.,* 1999). As a consequence, the introduction of validated prediction models could be a useful tool in selecting the optimum treatment regimens for subfertile couples, taking into account their individual pregnancy chances.

There are several clinical prediction models for the prediction of spontaneous pregnancy (Eimers *et al.,* 1994b; Wichmann *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997; Hunault *et al.,* 2004), pregnancy after intrauterine insemination (IUI) (Steures *et al.,* 2004) and pregnancy after IVF (Haan *et al.,* 1991; Templeton *et al.,* 1996; Stolwijk *et al.,* 1996), of which some have been validated (Stolwijk *et al.,* 1998; Hunault *et al.,* 2002; Hunault *et al.,* 2005). These models are intended to help gynaecologists in patient communication and decision making about the timing of treatment.

In this study, we aim to determine the concordance between gynaecologists to predict pregnancy chances, the concordance between gynaecologists in their treatment decisions and the influence of using prediction models on the concordance in treatment decisions.

Methods

We constructed 16 fictional vignettes of couples who underwent the basic fertility work up. All women had a regular cycle. The vignettes differed on the following seven prognostic factors: woman age, previous pregnancy, duration of subfertility, basal follicle-stimulating hormone (FSH) level, Chlamydia antibody titre (CAT), outcome of the postcoital test (PCT) and sperm motility. We selected these factors since they are the main prognosticators in existing prediction models in fertility, except for basal FSH level and CAT. The latter two tests have recently been introduced into the fertility work up and as such are likely to influence decision making.

Woman age was 25, 32 or 38 years. The duration of subfertility, defined as failure to conceive after one year of frequent unprotected intercourse, was one, two or three years. The basal FSH level on cycle day three was 6 or 15 IU/L. The CAT was 1:8 or 1:128. The outcome of the PCT was: 'no spermatozoa', 'motile, non-progressive spermatozoa' (motile spermatozoa that are not moving forwards) or 'progressive spermatozoa' (motile spermatozoa that are moving forwards). Progressive sperm motility was 15%, 35% or 65%. Tubal patency was confirmed by either hysterosalpingography or laparoscopy, depending on the result of the CAT.

Vignettes were generated from an orthogonal design. Orthogonal designs are constructed in such a way that inferences are based on main effects. Level combinations necessary for estimating second and higher order effects are excluded. Thereby the required number of measurements can be reduced (Addelman 1962). We used seven factors. Combining all factors would have resulted in 648 unique cases, whereas our orthogonal design needed only 16 cases without losing statistical information. Table 1a shows the composition of the 16 case vignettes.

Institutional Review Board approval was not requested, since no real patients were involved in this study.

Table 1a Overview of the 16 constructed vignettes of subfertile couples who underwent the basic fertility work-up (sorted by female age, duration of subfertility and type of subfertility, respectively).

Case no.	Woman age (y)	Duration of subfertility (y)	Subfertility	Basal FSH (IU/L)	CAT (titre)	PCT	Semen (% prog. sperm motility)
13	25	1	Primary	6	1:8	Motile, prog.	15
12	25	1	Primary	15	1:128	Motile, prog.	15
8	25	1	Secondary	6	1:128	Motile, nonprog.	65
1	25	1	Secondary	15	1:8	Motile, nonprog.	35
7	25	2	Primary	15	1:128	Nonmotile	65
15	25	2	Secondary	15	1:8	Motile, prog.	15
3	25	3	Primary	6	1:8	Nonmotile	35
10	25	3	Secondary	6	1:128	Motile, prog.	15
4	32	1	Primary	6	1:8	Motile, prog.	65
16	32	1	Secondary	15	1:8	Nonmotile	15
2	32	2	Secondary	6	1:128	Motile, prog.	35
5	32	3	Primary	15	1:128	Motile, nonprog.	15
14	38	1	Primary	15	1:128	Motile, prog.	35
9	38	1	Secondary	6	1:128	Nonmotile	15
11	38	2	Primary	6	1:8	Motile, nonprog.	15
6	38	3	Secondary	15	1:8	Motile, prog.	65

Prog. = progressive.

Thirty-five gynaecologists working in 14 academic and nonacademic hospitals participated in our survey. For all participants, we recorded the treatments they offered in their clinics, the time they had been practicing in fertility care, age and whether they used prediction rules in their daily care. Participants were asked for each of the 16 cases to appraise the one-year probability of spontaneous pregnancy, the probability to conceive after one IUI cycle and the probability to conceive after one IVF cycle (questionnaire 1). This estimation had to be marked on a scale, ranging from 0 to 100%, and divided into steps of 10%. After estimating the probabilities, they were asked to make a treatment decision: expectant management for at least six months, IUI for six cycles (either with or without ovarian hyperstimulation) or fresh IVF for three cycles (including frozen replacement cycles).

All participants received a cover letter together with the 16 cases, explaining how to use the prediction models for spontaneous pregnancy, IUI and IVF. In case of any remaining uncertainty, the participants could contact the first author.

Three months later, the participant gynaecologists again had to propose therapeutic regimes for the same 16 fictional cases but this time accompanied by pregnancy chances obtained from prediction models: prediction on spontaneous pregnancy, IUI and IVF (questionnaire 2) (Hunault *et al.,* 2004; Steures *et al.,* 2004; Templeton *et al.,* 1996).

Analysis

The concordance between the respondents for the first questionnaire was assessed in two ways. First, we analysed the concordance between gynaecologists of their probability estimates for spontaneous pregnancy, IUI and IVF by calculation of the interclass correlation coefficient (ICC) (Fleiss 1981). The ICC indicates the fraction of true variance from the total variance. Its value can vary between 0 and 1. The ICC can be considered as a measure of concordance for continuous variables.

Second, we analysed the concordance between the gynaecologists in their treatment decisions, by calculating Cohen's kappa (κ), a concordance statistic for categorical outcomes (Cohen 1960). Cohen's κ also takes values in the 0 to 1 range. A value of 1 would indicate perfect concordance whereas a value of 0 would indicates no concordance between the gynaecologists.

Finally, we analysed whether the predicted probabilities were associated with the proposed treatment regimens, with logistic regression analysis,

using a *p*-level for significance of 0.05. We assumed that one year of IUI treatment consists of six IUI cycles and one year of IVF treatment consists of three IVF cycles.

In the second questionnaire, the results of formal prediction models had been added to the case profiles. We assessed in what proportion the prediction models influenced concordance between gynaecologists in their treatment decisions, by comparing the concordance of treatment decisions before and after exposure to prediction models. The overall concordance in treatment decisions was tested with the McNemar statistic (*p*-level: 0.05), a nonparametric test for two related dichotomous variables.

Results

Of the 35 participating gynaecologists, 32 completed both questionnaires (91%). Of the respondents, 25 (71%) were working in specialised fertility units, and ten were working at general departments of obstetrics and gynaecology (29%). Their mean age was 40.2 years (minimum-maximum [min-max]: 26-57 years) and their mean experience in clinical practice was 8.9 years (min-max: 1.7 – 24 years). Twenty-eight respondents (80%) could offer IVF in their clinic and all respondents could offer IUI. Twenty-four participants (69%) reported working with formal prediction models for spontaneous pregnancy. None of the gynaecologists was used to work with prediction models for IUI or IVF.

The mean probability estimates for spontaneous pregnancy, IUI and IVF and the subsequent treatment decision are shown in table 1b.

Table 1b Overview of the 16 constructed vignettes of subfertile couples who underwent the basic fertility work-up (sorted by female age, duration of subfertility and type of subfertility, respectively). Details of cases are given in table 1a.

		Probability estimates		
Case no.	**N**	**Spontaneous pregnancy within 12 months, median (min–max)**	**Success rate of one IUI cycle, median (min–max)**	**Success rate of one IVF cycle, median (min–max)**
13	34	0.50 (0.20–0.90)	0.12 (0.03–0.40)	0.30 (0.20–0.68)
12	34	0.30 (0.10–0.90)	0.10 (0.02–0.63)	0.20 (0.10–0.45)
8	35	0.50 (0.25–0.90)	0.10 (0.03–0.40)	0.26 (0.10–0.85)
1	35	0.40 (0.15–0.90)	0.10 (0.05–0.40)	0.25 (0.10–0.60)
7	35	0.25 (0.10–0.90)	0.10 (0.02–0.40)	0.20 (0.08–0.60)
15	34	0.40 (0.05–0.90)	0.10 (0.03–0.40)	0.22 (0.05–0.50)
3	33	0.20 (0.10–0.50)	0.10 (0.03–0.60)	0.25 (0.15–0.60)
10	35	0.25 (0.03–0.90)	0.10 (0.03–0.40)	0.25 (0.10–0.65)
4	33	0.50 (0.20–0.90)	0.12 (0.05–0.55)	0.25 (0.15–0.60)
16	34	0.30 (0.06–0.90)	0.10 (0.03–0.40)	0.20 (0.06–0.60)
2	35	0.30 (0.10–0.70)	0.10 (0.04–0.55)	0.25 (0.15–0.65)
5	35	0.10 (0.03–0.50)	0.10 (0.01–0.30)	0.20 (0.08–0.60)
14	34	0.20 (0.05–0.90)	0.09 (0.02–0.40)	0.19 (0.05–0.40)
9	35	0.20 (0.05–0.80)	0.10 (0.03–0.40)	0.20 (0.10–0.40)
11	34	0.20 (0.04–0.65)	0.10 (0.03–0.30)	0.20 (0.10–0.40)
6	35	0.18 (0.03–0.50)	0.08 (0.00–0.30)	0.18 (0.05–0.30)
Total	550 (98%)			

N = number of respondents per case.

The column 'probability estimates' shows, per case, the median probability to conceive spontaneously, to conceive after IUI and to conceive after IVF as estimated by 35 gynaecologists. Ranges show the minimum and maximum (min-max) probability estimates.

Table 1c Overview of the 16 constructed vignettes of subfertile couples who underwent the basic fertility work-up (sorted by female age, duration of subfertility and type of subfertility, respectively). Details of cases are given in table 1a.

		Overall treatment decision		
Case no.	**N**	**Expectant management (6 months) (%)**	**IUI (%)**	**IVF (%)**
13	34	88.2	11.8	
12	33	39.4	36.4	24.2
8	33	81.8	15.2	3.0
1	35	71.4	14.3	14.3
7	34	23.5	50.0	26.5
15	34	41.2	41.2	17.6
3	35	5.7	82.9	11.4
10	34	11.8	70.6	17.6
4	33	97.0	3.0	
16	34	32.4	50.0	17.6
2	32	62.5	28.1	9.4
5	32	3.1	56.3	40.6
14	33	18.2	30.3	51.5
9	33	30.3	42.4	27.3
11	34	2.9	61.8	35.3
6	35	5.7	37.1	57.1
Total	538 (96%)			

N = number of respondents per case.

The column 'overall treatment decision' shows the proportion of gynaecologists who chose for a specific treatment per case. In this table, cases were sorted by woman age, duration of subfertility and basal FSH level.

Concordance between gynaecologists was poor for the probability estimates of spontaneous pregnancy, pregnancy after IUI and pregnancy after IVF, with ICCs varying between 0.05 and 0.34 (table 2). In addition, concordance between gynaecologists with respect to their treatment decisions based on their intuitive probability estimates was also poor (table 2).

Table 2 Concordance between gynaecologists with respect to probability estimates of pregnancy rates and treatment decisions.

	Questionnaire 1 (based on intuition)		Questionnaire 2 (based on formal prediction rules)		Difference between questionnaire 2 and questionnaire 1	
		95% CI		95% CI		95% CI
Concordance of probability estimates for spontaneous pregnancy (ICC)	0.34	0.17-0.69	–	–	–	–
Concordance of the probability estimates for pregnancy after IUI (ICC)	0.05	0.01-0.19	–	–	–	–
Concordance of the probability estimates for pregnancy after IVF (ICC)	0.15	0.06-0.43	–	–	–	–
Concordance of treatment decisions between gynaecologists (κ)	0.21	0.19-0.24	0.38	0.34-0.43	0.17	0.12-0.22

ICC = interclass correlation coefficient; κ = Cohen's kappa; CI = confidence interval.

Overall, the treatment decisions made by gynaecologists were consistent with the ranking of their probability estimates. In the univariable analysis, the choice for expectant management was significantly influenced by the probability estimates of spontaneous pregnancy, IUI and IVF, the difference between probabilities of spontaneous pregnancy and IUI and the difference between the probabilities of spontaneous pregnancy and IVF (table 3).

In the multivariable analysis, only the probability of spontaneous pregnancy was a statistically significant predictor for the choice for expectant management and IUI (OR: 3.5; 95% CI: 2.8 – 4.3). No other factor was significant once this factor had entered the model.

Table 3 Association of intuitive probability estimates and the subsequent treatment decision (i.e. expectant management for 6 months); results of the univariable and multivariable regression analyses.

	Univariable analysis			**Multivariable analysis**		
	OR	**95% CI**	***p*-value**	**OR**	**95% CI**	***p*-value**
Intuitive probability for (per 10% increase)						
Spontaneous pregnancy (within 12 months)	3.5	2.8-4.3	< 0.001	3.5	2.8-4.3	< 0.001
Pregnancy with IUI (six cycles)	1.2	1.1-1.3	< 0.001	-		
Pregnancy with IVF (three cycles)	1.6	1.4-1.9	< 0.001	-		
Difference between probability estimates of						
Spontaneous pregnancy versus IUI pregnancy	1.6	1.5-1.8	< 0.001	-		
Spontaneous pregnancy versus IVF pregnancy	2.0	1.7-2.3	< 0.001	-		

OR = odds ratio; CI = confidence interval.

The results of the second questionnaire are shown in table 2. After prediction models had been added to the case profiles, the concordance

between the gynaecologists with respect to their treatments decisions increased from 0.21 to 0.38 (difference: 0.17, 95% CI: 0.12 – 0.22).
The impact of the use of prediction models on the treatment decision by gynaecologists is shown in table 4. Due to missing data, we could analyse 495 out of 560 treatment decisions (89%). Of the 192 initial proposals for expectant management, 24 (13%) were changed to IUI and five (3%) to IVF. Of the 188 treatment proposals for IUI, 66 (35%) were changed to expectant management and eleven (6%) to IVF. Of the 115 treatment proposals for IVF treatment, 25 (22%) were changed to expectant management and 39 (34%) to IUI.

Overall, the percentage of expectant management increased from 39% (192 out of 495) to 51% (254 out of 495) after prediction models had been added. Consequently, the percentage IUI dropped from 38% (188 out of 495) to 35% (174 out of 495) and the percentage of IVF treatment dropped from 23% (115 out of 495) to 14% (67 out of 495). This overall change from 61% treatment to 49% treatment was statistically significant ($p < 0.001$, McNemar's test).

Table 4 Comparison of treatment decisions made by the participating gynaecologists in questionnaire 1 and questionnaire 2.

		Treatment decisions in questionnaire 2 (with prediction rules)			
		Expectant man.	**IUI**	**IVF**	**Total**
Treatment decisions in questionnaire 1 (without prediction rules)	**Expectant management**	163	24	5	192
	IUI	66	111	11	188
	IVF	25	39	51	115
	Total	254	174	67	495

Discussion

The importance of estimating conception chances after completion of the fertility work up has been stressed before (Gnoth *et al.,* 2005; Te Velde and Cohlen 1999). The key question of whether gynaecologists are able to do this has never been addressed and, therefore, this study aimed to determine the concordance between gynaecologists in predicting pregnancy chance and in treatment decisions and to assess in what proportion prediction models influenced this concordance.

Unfortunately, the gynaecologists differed widely in estimating pregnancy chances and there was a huge variation in the proposed therapeutic regimens. However, the treatment decisions made by gynaecologists were consistent with the ranking of their probability estimates. When prediction models were used, the concordance for treatment decisions increased only slightly. With prediction models, more gynaecologists were likely to propose expectant management in favour of IVF.

Why did the use of prediction models improve the concordance only slightly? The most obvious explanation may be the difference in interpretation of pregnancy chances. Whereas one doctor may feel that expectant management is still justified in case a 30% pregnancy chance without treatment is expected to increase to 60% with IVF, another doctor might recommend IVF treatment is this situation. This difference was seen when treatment decisions of older gynaecologists were compared with those of the younger ones. Gynaecologists older than 40 years (N = 18) were slightly more inclined to adopt a conservative approach (55% in all cases), as compared with gynaecologists younger than 40 years, who proposed expectant management in 48% of the cases.

Another explanation may be that gynaecologists may variably offer treatment despite a reasonable prognosis for expectant management in older women in view of the decreasing time remaining for conception, particularly if the couple are aiming for more than one child. This hypothesis was supported by the fact that 71% of the gynaecologists would advise treatment in older women (38 years) with a good prognosis (> 30% in 1 year), whereas only 16% would do so in younger women (25 and 32 years) with a good prognosis. Woman age has been reported to play an important role in clinical decision making in subfertility (Van der Steeg *et al.,* 2006).

In daily practice, the choice of treatment is a decision made by both the doctor and patient and depends on more variables than fecundity alone.

For example, in several countries, fertility treatment is not reimbursed, and the choice for treatment strongly depends on the economic situation of the patient (Jain *et al.,* 2002).

Some of the participating gynaecologists already used to work with prediction models in their clinic. The probability estimates of these gynaecologists may have been influenced by previous experience with the prediction models. Most clinicians only used prediction models for spontaneous pregnancy, not for IUI and IVF. This can be an explanation for the fact that concordance in prediction was better for spontaneous pregnancy than for IUI and IVF.

Although the PCT is considered to be a controversial test (Eimers *et al.,* 1994a; Oei *et al.,* 2001; Oei *et al.,* 1998; Oei 1998), it is used in the majority of the prediction models that are available (Eimers *et al.,* 1994b; Snick *et al.,* 1997; Hunault *et al.,* 2004). For this reason only, we chose to use the PCT in the vignettes of the present study. However, we do not believe that omission of the PCT from the paper cases would affect the results and conclusions of the present study.

The use of prediction models improved the concordance in treatment decisions from fair (κ 0.21) to moderate (κ 0.38). The clinical implication of this improvement can be explained in the following example. Let us assume that two gynaecologists have to counsel 100 subfertile couples. When the concordance between their treatment decisions is 0.2 and the fraction of decisions on expectant management, IUI and IVF are 39%, 38% and 23%, respectively, they will agree in treatment decisions in 48 couples. An increase of the concordance to a κ-value of 0.4 will implicate that treatment decisions will become equal in 61 couples.

One previous study on this topic reported a good concordance for the estimation of the probability of spontaneous conception (ICC = 0.71). This contrasts with the moderate concordance between respondents' probability estimations for spontaneous pregnancy found in our study (ICC = 0.34) (Wiegerinck *et al.,* 1999). This difference can be explained by the fact that in the previous study only four cases were used, of which two cases had rather extreme profiles. In contrast, the concordance of probability estimates for IVF was poor in the previous study and this study (ICC 0.24 and 0.15, respectively). The previous study did not assess the final treatment decision.

Gynaecologists favoured expectant management rather than IVF once prediction models were made available. An explanation may be that gynaecologists were more comfortable with expectant management once they could rely on predicted pregnancy rates rather than on their own

intuitive estimates. If this were to be the case, prediction models could not only reduce treatment differences between gynaecologists but also prevent treatment at an early stage of subfertility when prospects for spontaneous pregnancy are still good.

In conclusion, our study shows that gynaecologists differ widely in estimating prognosis in fictional cases of subfertile couples. Their proposed therapeutic regimens likewise show huge variation. The use of prediction models for spontaneous pregnancy, IUI and IVF in itself are thus not sufficient to guarantee uniform counselling of subfertile couples. Future studies designed to evaluate the potential effectiveness of the use of prediction models should therefore offer guidelines for clinicians on how to deal with these models.

Acknowledgements

The authors thank all participating gynaecologists for their contribution to this study. Furthermore, M Pel is acknowledged for her critical reading and valuable suggestions on previous drafts of this study (Department of Obstetrics and Gynaecology, Academic Medical Centre, Amsterdam, The Netherlands).

References

Addelman S. (1962) Symmetrical and asymmetrical fractional factorial plans. *Technometrics* 4,47-58.

Cohen J. (1960) A coefficient of agreement for nominal scales. *Educ Psych Meas* 20,46.

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Eimers JM, Te Velde ER, Gerritse R, Van Kooy RJ, Kremer J, and Habbema JDF. (1994a) The validity of the postcoital test for estimating the probability of conceiving. *Am J Obstet Gynecol* 171,65-70.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994b) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Fleiss,J.L. (1981) Statistical Methods for Rates and Proportions. John Wiley, New York.

Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, and Freundl G. (2005) Definition and prevalence of subfertility and infertility. *Hum Reprod* 20,1144-1147.

Haan G, Bernardus RE, Hollanders JM, Leerentveld RA, Prak FM, and Naaktgeboren N. (1991) Results of IVF from a prospective multicentre study. *Hum Reprod* 6,805-810.

Hunault CC, Eijkemans MJC, Te Velde ER, Collins JA, and Habbema JDF. (2002) Validation of a model predicting spontaneous pregnancy among subfertile untreated couples. *Fertil Steril* 78,500-506.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Hunault CC, Laven JS, Van Rooij IA, Eijkemans MJC, Te Velde ER, and Habbema JDF. (2005) Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum Reprod* 20,1636-1641.

Jain T, Harlow BL, and Hornstein MD. (2002) Insurance coverage and outcomes of in vitro fertilization. *N Engl J Med* 347,661-666.

Oei SG. (1998) The postcoital test: a controversial investigation. *Eur J Obstet Gynecol Reprod Biol* 77,123-124.

Oei SG, Helmerhorst FM, Bloemenkamp KW, Hollants FA, Meerpoel DE, and Keirse MJ. (1998) Effectiveness of the postcoital test: randomised controlled trial. *BMJ* 317,502-505.

Oei SG, Helmerhorst FM, and Keirse MJ. (2001) Routine postcoital testing is unnecessary. *Hum Reprod* 16,1051-1053.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Steures P, Van der Steeg JW, Mol BWJ, Eijkemans MJC, Van der Veen F, Habbema JDF, Hompes PGA, Bossuyt PMM, Verhoeve HR, Van Kasteren YM *et al*. (2004) Prediction of an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 82,45-51.

Stolwijk AM, Straatman H, Zielhuis GA, Jansen CA, Braat DD, Van Dop PA, and Verbeek AL. (1998) External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. *Hum Reprod* 13,3542-3549.

Stolwijk AM, Zielhuis GA, Hamilton CJ, Straatman H, Hollanders JM, Goverde HJ, Van Dop PA, and Verbeek AL. (1996) Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. *Hum Reprod* 11,2298-2303.

Te Velde ER and Cohlen BJ. (1999) The management of infertility. *N Engl J Med* 340,224-226.

Templeton A, Morris JK, and Parslow W. (1996) Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 348,1402-1406.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Bossuyt PMM, Hompes PGA, Van der Veen F, and Mol BWJ. (2006) Which factors play a role in clinical decision-making in subfertility? *Reprod Biomed Online*.

Wichmann L, Isola J, and Tuohimaa P. (1994) Prognostic variables in predicting pregnancy. A prospective follow up study of 907 couples with an infertility problem. *Hum Reprod* 9,1102-1108.

Wiegerinck MA, Bongers MY, Mol BWJ, and Heineman MJ. (1999) How concordant are the estimated rates of natural conception and in-vitro fertilization/embryo transfer success? *Hum Reprod* 14,689-693.

REPRODUCTIVE BIOMEDICINE ONLINE
2006; 12: 473 - 480

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Patrick M.M. Bossuyt
Peter G.A. Hompes
Fulco van der Veen
Ben Willem J. Mol

chapter (04)

WHICH FACTORS PLAY A ROLE
in clinical decision-making in subfertility?

Abstract

We determined the importance of factors of the fertility work-up in the subsequent management of subfertile patients.
Sixteen vignettes of subfertile couples were constructed by varying fertility history, post-coital test, sperm motility, FSH concentration and Chlamydia antibody titre (CAT). Thirty-five gynaecologists estimated probabilities of treatment-independent pregnancy, intrauterine insemination (IUI) and IVF. Thereafter, they chose IUI, IVF or no treatment. The relative contribution of each factor to probability estimates and to subsequent treatment decisions was calculated.
Duration of subfertility and maternal age were the most important contributors for gynaecologists' estimates of treatment-independent pregnancy (RC: 41%, 26%). Maternal age and FSH concentration were the most important contributors in the estimates for IUI (relative contribution (RC): 51%, 25%) and for IVF (RC: 64%, 31%). The decision to start IVF was mainly determined by maternal age, duration of subfertility, FSH concentration and CAT.
The relative contribution of maternal age and duration of subfertility was in concordance with existing prediction models, whereas previous pregnancy and FSH concentration were under- and overestimated respectively.
In conclusion, maternal age, duration of subfertility and FSH concentration are the main factors in clinical decision-making in subfertility. Gynaecologists overestimate the importance of FSH concentration, but underestimate that of a previous pregnancy, as compared with their importance reported in prediction models and guidelines.

Introduction

In many countries, there is consensus about which tests should be performed in the basic fertility work-up. According to the National Institute for Excellence (NICE) guidelines, both semen analysis and cycle assessment should be performed, and the tubal status should be assessed with a hysterosalpingogram (HSG) (National institute for clinical excellence 2004). In the guidelines of the Dutch society of Obstetrics and Gynaecology, similar tests are recommended (Dutch Society of Obstetrics and Gynaecology 2004). However, tubal status can also initially be assessed with a *Chlamydia* antibody titre (CAT) (Mol *et al.,* 1997; Akande *et al.,* 2003). The post-coital test (PCT) is optional. Maternal age, duration and subfertility and fertility, being primary or secondary, are mentioned explicitly as parts of the medical history. Nevertheless, performance of the basic fertility work-up according to guidelines does not imply that information obtained with these tests is used in subsequent decision-making in a correct manner.

The prognostic capacity of most factors of the basic fertility work-up has been assessed in large cohort studies, and integrated in prediction models for treatment-independent pregnancy, intrauterine insemination (IUI) and IVF (Eimers *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997; Hunault *et al.,* 2004; Steures *et al.,* 2004; Templeton *et al.,* 1996).

However, at present, little is known how individual gynaecologists weigh the information of each individual factor of the fertility work-up in their actual management decisions. In view of this lack of knowledge, this study investigated which factors of the basic fertility work-up are of importance for gynaecologists both in their estimates of pregnancy chances of treatment-independent pregnancy, IUI and IVF, and in their subsequent management decisions. The importance of each of these factors as found in this study was also compared with their importance as reported in existing prediction models and guidelines.

Materials and methods

Sixteen fictional vignettes were constructed of couples who underwent the basic fertility work-up. All women had a regular cycle. The vignettes differed on the following seven prognostic factors: maternal age, previous pregnancy, duration of subfertility, basal FSH concentration, CAT, outcome of the PCT and sperm motility. These factors were selected because they are the main prognosticators in existing prediction models in fertility, except for basal FSH concentration and CAT. The latter two tests have recently been introduced into the fertility work-up and as such are likely to influence decision-making.

Maternal age was 25, 32 or 38 years. The duration of subfertility, defined as failure to conceive after one year of frequent unprotected intercourse, was one, two or three years. The basal FSH concentration on cycle day three was 6 or 15 IU/L. The CAT was 1:8 or 1:128. The outcome of the PCT was: 'no spermatozoa', 'motile, nonprogressive spermatozoa' (motile spermatozoa that are not moving forwards) or 'progressive spermatozoa' (motile spermatozoa that are moving forwards). Progressive sperm motility was 15%, 35% or 65%. Tubal patency was confirmed by either hysterosalpingography (HSG) or laparoscopy, depending on the result of the CAT.

Vignettes were generated from an orthogonal design. Orthogonal designs are constructed in such a way that inferences are based on main effects. Level combinations necessary for estimating second and higher order effects are excluded. Thereby the required number of measurements can be reduced (Addelman 1962). Seven factors were used. Combining all factors would have resulted in 648 unique cases, whereas our orthogonal design needed only 16 cases without losing statistical information. Table 1 shows the composition of the 16 case vignettes and the fertility profiles.

Table 1 Overview of the characteristics of 16 constructed vignettes of subfertile couples who underwent the basic fertility work-up (sorted by maternal age, duration of subfertility and type of subfertility respectively).

Case no.	Maternal age (y)	Duration of subfertility (y)	Subfer-tility	Basal FSH (U/L)	CAT (titre)	PCT	Semen % prog motile
13	25	1	Primary	6	1:8	Motile, prog.	15
12	25	1	Primary	15	1:128	Motile, prog.	15
8	25	1	Secon-dary	6	1:128	Motile, non-prog.	65
1	25	1	Secon-dary	15	1:8	Motile, non-prog.	35
7	25	2	Primary	15	1:128	Non-motile	65
15	25	2	Secon-dary	15	1:8	Motile, prog.	15
3	25	3	Primary	6	1:8	Non-motile	35
10	25	3	Secon-dary	6	1:128	Motile, prog.	15
4	32	1	Primary	6	1:8	Motile, prog.	65
16	32	1	Secon-dary	15	1:8	Non-motile	15
2	32	2	Secon-dary	6	1:128	Motile, prog.	35
5	32	3	Primary	15	1:128	Motile, non-prog.	15
14	38	1	Primary	15	1:128	Motile, prog.	35
9	38	1	Secon-dary	6	1:128	Non-motile	15
11	38	2	Primary	6	1:8	Motile, non-prog.	15
6	38	3	Secon-dary	15	1:8	Motile, prog.	65

CAT = *Chlamydia* antibody titre; PCT = postcoital test; prog. = progressive.

Thirty-five gynaecologists, working in 14 academic and non-academic hospitals, participated in our survey. For all participants, the treatments offered in their clinics, the time they had been practicing in fertility care, and whether they used prediction rules in their daily care were recorded. Participants were asked for each of the 16 cases to appraise the one-year probability of treatment-independent pregnancy, the probability of conceiving after one IUI cycle and the probability of conceiving after one IVF cycle. This estimation had to be marked on a scale ranging from 0% to 100%, and divided into steps of 10%. After estimating the probabilities, participants were asked to make a treatment decision: expectant management for at least six months, IUI for six cycles (either with or without ovarian hyperstimulation) or fresh IVF for three cycles (including frozen replacement cycles).

In summary, the 35 gynaecologists were asked to estimate in total 1680 pregnancy probabilities (35 gynaecologists x 16 cases x 3 probabilities) and in total 560 treatment decisions. It was assumed that this number represents a large enough sample size to obtain reliable and powered outcomes.

Analysis

The relative contribution (RC) of each factor to the probability estimates for treatment-independent pregnancy, IUI and IVF was calculated, in order to assess which factors from the fertility work-up are involved in gynaecologists' estimates of success rates of fertility treatment. The RC was estimated with multivariable linear regression analysis (enter method; *p*-level 0.05). The RC of each factor was calculated as the proportion of the squared partial correlation over the sum of squares of partial correlations of all seven factors. The RC expresses the contribution of each factor to the estimate of the pregnancy chance as made by the clinician. The sum of the RC always adds up to 100%. Factors analysed were the seven factors that were varied over the vignettes; maternal age, duration of subfertility, subfertility being primary or secondary, semen analysis, PCT, basal FSH concentration and the CAT. The estimated probability of pregnancy (log odds) was the dependent variable.

The association of each factor and the final treatment decision were assessed, in order to assess which factors were of importance for these subsequent treatment decisions. Since choice of treatment was categorical (expectant management, IUI and IVF), odds ratios (OR) and corresponding 95% confidence intervals (CI) were calculated. The OR were calculated

using logistic regression analyses (forward stepwise; *p*-level 0.05). Two analyses were performed. In the first analysis, the decision to start treatment (either IUI or IVF) was compared with the decision for expectant management (i.e. no treatment). In the second analysis, the specific choice for IVF was compared with no IVF (expectant management or IUI). In cases where the OR was > 1, gynaecologists tended to select a treatment option, whereas an OR < 1 indicated that gynaecologists were less likely to choose treatment.

Finally, the RC of prognostic factors as analysed in this study was compared with data from prediction models on treatment-independent pregnancy, on pregnancy after IUI and on pregnancy after IVF (Hunault *et al.,* 2004; Steures *et al.,* 2004; Templeton *et al.,* 1996). Since basal FSH concentration was not assessed in these prognostic models, data on basal FSH concentration were obtained from a cohort study on prediction of treatment-independent pregnancy, and from a meta-analysis on the prediction of IVF (Van Montfrans et al., 2000; Bancsi et al., 2000; Bancsi et al., 2003).

The factors integrated in the prediction models were classified using a semi quantitative approach, in which a *p*-level ≤ 0.05 was classified as very important (++) a *p*-level of 0.05-0.30 as important (+) and a *p*-level of > 0.30 as not important (-).

Results

Twenty-five of the 35 gynaecologists (71%) were working in specialized fertility units, whereas 10 gynaecologists (29%) were working in general departments of obstetrics and gynaecology. Twenty-eight gynaecologists (80%) could offer IVF in their clinic, and all gynaecologists could offer IUI. The mean time that gynaecologists had been working in clinical practice was 8.9 years (range 1.7 – 24 years). All gynaecologists performed the basic fertility work-up according to the guidelines of the Dutch Society of Obstetrics and Gynaecology. Twenty-four gynaecologists (69%) reported the use of formal prediction rules for treatment-independent pregnancy. None of the gynaecologists was used to working with prediction rules for IUI or IVF.

Table 2 and 3 show the estimated probability of spontaneous pregnancy, pregnancy after IUI and after IVF and the percentage of gynaecologists opting subsequently for expectant management, IUI or IVF for each of the sixteen cases. For example, in case 1, the median estimated probability of all 35 estimates given by the gynaecologists to conceive independently of treatment within 12 months was 0.40 (40% chance), with a range from 0.15 to 0.90. The median success rate of IUI per cycle was estimated to be 0.10 (10% chance) and the median success rate of IVF to be 0.25 (25% chance). Overall, 72% of the 35 gynaecologists opted for expectant management in this case, whereas 14% decided to start IUI and 14% decided to start IVF.

Table 2 Probability estimates. The table shows, per case, the median probability to conceive independently of treatment, to conceive after IUI and to conceive after IVF as estimated by 35 gynaecologists. Ranges show the minimum and maximum probability estimates. Details of cases are given in table 1.

Case no.	N	Spontaneous pregnancy within 12 months Median (min–max)	Success rate of one IUI cycle Median (min–max)	Success rate of one IVF cycle Median (min–max)
13	34	0.50 (0.20–0.90)	0.12 (0.03–0.40)	0.30 (0.20–0.68)
12	34	0.30 (0.10–0.90)	0.10 (0.02–0.63)	0.20 (0.10–0.45)
8	35	0.50 (0.25–0.90)	0.10 (0.03–0.40)	0.26 (0.10–0.85)
1	35	0.40 (0.15–0.90)	0.10 (0.05–0.40)	0.25 (0.10–0.60)
7	35	0.25 (0.10–0.90)	0.10 (0.02–0.40)	0.20 (0.08–0.60)
15	34	0.40 (0.05–0.90)	0.10 (0.03–0.40)	0.22 (0.05–0.50)
3	33	0.20 (0.10–0.50)	0.10 (0.03–0.60)	0.25 (0.15–0.60)
10	35	0.25 (0.03–0.90)	0.10 (0.03–0.40)	0.25 (0.10–0.65)
4	33	0.50 (0.20–0.90)	0.12 (0.05–0.55)	0.25 (0.15–0.60)
16	34	0.30 (0.06–0.90)	0.10 (0.03–0.40)	0.20 (0.06–0.60)
2	35	0.30 (0.10–0.70)	0.10 (0.04–0.55)	0.25 (0.15–0.65)
5	35	0.10 (0.03–0.50)	0.10 (0.01–0.30)	0.20 (0.08–0.60)
14	34	0.20 (0.05–0.90)	0.09 (0.02–0.40)	0.19 (0.05–0.40)
9	35	0.20 (0.05–0.80)	0.10 (0.03–0.40)	0.20 (0.10–0.40)
11	34	0.20 (0.04–0.65)	0.10 (0.03–0.30)	0.20 (0.10–0.40)
6	35	0.18 (0.03–0.50)	0.08 (0.00–0.30)	0.18 (0.05–0.30)
Total	550 (98%)			

IUI = intrauterine insemination; N = number of respondents per case.

Table 3 Overall treatment decision. The table shows the proportion of gynaecologists who chose for a specific treatment per case. In this table cases were sorted by maternal age, duration of subfertility and basal FSH concentration. Details of cases are given in table 1.

Case no.	N	Expectant management (6 months) (%)	IUI (%)	IVF (%)
13	34	88.2	11.8	
12	33	39.4	36.4	24.2
8	33	81.8	15.2	3.0
1	35	71.4	14.3	14.3
7	34	23.5	50.0	26.5
15	34	41.2	41.2	17.6
3	35	5.7	82.9	11.4
10	34	11.8	70.6	17.6
4	33	97.0	3.0	
16	34	32.4	50.0	17.6
2	32	62.5	28.1	9.4
5	32	3.1	56.3	40.6
14	33	18.2	30.3	51.5
9	33	30.3	42.4	27.3
11	34	2.9	61.8	35.3
6	35	5.7	37.1	57.1
Total	538 (96%)			

IUI = intrauterine insemination; N = number of respondents per case.

Table 4 shows the impact of the seven factors for the prediction of treatment-independent pregnancy, pregnancy after IUI and pregnancy after IVF made by the gynaecologists. Duration of subfertility (RC of 41%) and maternal age (RC of 26%) were the most important factors for the prediction of treatment-independent pregnancy by a gynaecologist. The RC of the PCT, semen analysis, CAT, basal FSH concentration and subfertility being primary or secondary were limited.

Table 4 Impact of seven factors of the basic fertility work-up to gynaecologists' prediction of treatment-independent pregnancy, IUI and IVF, expressed as relative contribution. Data are relative contribution (RC in %), representing the weight of each factor in the estimation of prognosis.

	Prediction of treatment-independent pregnancy (within 12 months) (%)	**Prediction of pregnancy after IUI (per cycle) (%)**	**Prediction of pregnancy after IVF (per cycle) (%)**
Duration of subfertility (y)	41	13	4.7
Maternal age (y)	26	51	64
Post-coital test	9.6	0.4	0.2
Basal FSH-level (IU/L)	8.5	25	31
Semen motility (% progressive motile)	7.1	3.3	0.2
CAT (titre)	5.4	7.8	0.3
Subfertility (primary or secondary)	2.4	0.1	0.0
Total	100	100	100

CAT = *Chlamydia* antibody titre; IUI = intrauterine insemination.

Maternal age was the most important factor for the prediction of pregnancy after IUI as estimated by the gynaecologist (RC of 51%), followed by basal FSH concentration and duration of subfertility (RC of 25 and 13% respectively). The RC of the other factors for IUI was of limited importance. Maternal age and basal FSH concentration were also the most important factors for the prediction of pregnancy after IVF as estimated by the gynaecologist (RC of 64 and 31% respectively), with a limited contribution of all other factors.

Table 5 shows the impact of the seven factors from the basic fertility work-up on gynaecologists' treatment decisions. Gynaecologists were more likely to opt for treatment in cases of higher maternal age (38 and

32 years), longer duration of subfertility (3 and 2 years), primary subfertility, low semen motility (35 and 15%), an abnormal PCT (nonprogressive and non motile), elevated basal FSH concentration (15 IU/L) and a positive CAT result (1:128).

Gynaecologists were more likely to choose IVF instead of no IVF (IUI or expectant management) in cases of higher maternal age (38 years), longer duration of subfertility (3 and 2 years), elevated basal FSH concentration (15 IU/L) and positive CAT result (1:128).

Table 5 Impact of the seven factors from the basic fertility work-up on gynaecologists' treatment decisions.

		Treatment (IUI or IVF) versus expectant management		**IVF versus no IVF (expectant management or IUI)**	
Factor		**OR**	**95% CI**	**OR**	**95% CI**
Maternal age	25	1		1	
(y)	32	1.7	(0.8-3.7)	0.8	(0.4.7)
	38	5.9	(3.0-11)	5.6	(3.3-9.5)
Duration of	1	1		1	
subfertility	2	4.6	(2.5-7.7)	2.2	(1.1-4.3)
(y)	3	50	(17-100)	3.4	(1.7-6.6)
Subfertility	Secondary	1		1	
	Primary	2.1	(1.2-3.6)	0.9	(0.6.6)
Post-coital test	Progressive	1		1	
(motility)	Motile, non-progressive	3.7	(1.4-10)	1.3	(0.7-2.5)
	Non-motile	3.0	(1.7-5.3)	1.5	(0.8-2.8)
Semen	65	1		1	
(% progressive	35	2.2	(1.0-4.6)	0.90	(0.5-1.6)
motile)	15	4.8	(2.2-10)	0.5	(0.2-1.0)
Basal FSH	6	1		1	
concentration	15	3.9	(2.2-6.3)	4.5	(2.5-8.2)
(IU/L)					
CAT (titre)	1:8	1		1	
	1:128	2.1	(1.2-3.6)	1.8	(1.0-3.0)

CAT = *Chlamydia* antibody titre; CI = confidence interval; IUI = intrauterine insemination; OR = odds ratio.

Table 6 shows the comparison of the importance of the factors used in actual clinical decision-making and their value as reported in current prognostic models. Maternal age and duration of subfertility were important for the prediction of treatment-independent pregnancy in current models. Gynaecologists indeed used these two factors for their estimates of treatment-independent pregnancy. In contrast, the factors subfertility being primary or secondary, the PCT and semen analysis were all important in the prediction model, but were of limited importance in the actual pregnancy estimates made by these gynaecologists.

For the prediction of pregnancy after IUI, maternal age, basal FSH concentration and duration of subfertility played an important role in the process of decision-making by these gynaecologists, whereas there was a small impact of the semen analysis and the CAT. This was in accordance with the results of the existing prediction models.

For the prediction of pregnancy after IVF, maternal age was important in the IVF prediction model. In concordance, gynaecologists used maternal age in their own intuitive IVF estimates. In contrast, the factors duration of subfertility and subfertility being primary or secondary were important in the prediction model, but were of limited importance in the actual prediction making by these gynaecologists. In contrast to the literature, where the role of FSH is not important in the prediction of pregnancy after IVF, basal FSH concentration was found to play an important role in the process of decision-making by gynaecologist in this study.

Table 6 Semi-quantitative comparison of the importance of factors for actual decision making as found in this study, compared with their prognostic value as reported in existing prediction models for treatment-independent pregnancy (Hunault *et al.*), IUI (Steures *et al.*) and IVF (Templeton *et al.*).

	Prediction of treatment-independent pregnancy		**Prediction of IUI**		**Prediction of IVF**	
	Model of Hunault	**This study**	**Model of Steures**	**This study**	**Model of Templeton**	**This study**
Maternal age	+ +	+ +	+ +	+ +	+ +	+ +
Duration of subfertility	+ +	+ +	+	+ +	+ +	+
Subfertility (primary/ secondary)	+ +	–	–	–	+ +	–
Post-coital test	+ +	+	+ +	–	n.a.	–
Semen analysis	+ +	+	+	+	n.a.	–
Basal FSH concentration	– [b]	+	n.a.	+ +	– [b]	+ +
Tubal status[a]	n.a.	+	+	+	+ +	–

[a] In this study tubal status is tested with a *Chlamydia* antibody titre and depending on the result subsequently confirmed with hysterosalpingography or laparoscopy.

[b] Prognostic value of basal FSH concentration as reported on in literature.

Indications of importance:

+ + = very important (*p*-value 0-0.05);

+ = important (*p*-value 0.05-0.3);

– = not important (*p*-value > 0.3);

n.a.= not available as factor in existing prediction models.

Discussion

This study determined the importance of each individual factor of the fertility work-up in gynaecologists' clinical decision-making. The study was performed among gynaecologists working in The Netherlands. Maternal age and duration of subfertility contributed for over 60% of the estimation of the probability of treatment-independent pregnancy. Maternal age and basal FSH were the two most important contributors in the estimation of pregnancy chances after IUI and IVF. Duration of subfertility was less important. All factors were important in the subsequent decision for either treatment or expectant management. Maternal age, duration of subfertility, basal FSH and CAT mainly determined the specific decision for IVF, the other factors playing a statistically insignificant role.

This study did not analyse implementation of the fertility work-up, but the subsequent step after implementation. In other words, the study focused on whether results of the fertility tests contributed to the clinical decision-making by gynaecologists. On the one hand, the study analysed how important these results were in the estimation of pregnancy chance made by gynaecologists. On the other hand, it analysed in which proportion these results were involved in clinical decision-making. So far as is known, such a study has not been performed before in reproductive medicine.

A weakness of this study is the use of paper cases. Artificially constructed paper cases cannot substitute for real doctors dealing face to face with subfertile patients in clinical circumstances. However, this study can be seen as a first step in the assessment of how gynaecologists use information of fertility work-up in subsequent clinical decision-making.

A methodological issue is that the orthogonal design does not reflect the prevalence of patient characteristics in reality. For example, in our orthogonal design the basal FSH concentration was elevated in 50% of the cases, whereas this will be the case in less than 10% of subfertile women in reality. As a consequence, the relative contribution of some factors might change in true clinical circumstances, since the relative contribution depends on the prevalence of a specific test result. For basal FSH, this has not resulted in discrepancy. With respect to prediction of treatment-independent pregnancy, the findings were in concordance with current literature, which reported basal FSH of limited predictive capacity (Van Rooij *et al.,* 2006). In this study, its relative contribution was < 10%.

The data also show that the contribution of maternal age and the contribution of duration of subfertility to the process of clinical decision-

making are in accordance with the results of existing prediction models. In contrast, the use of information on previous pregnancies or basal FSH concentration was not. Gynaecologists did not use information on previous pregnancies, whereas this factor has been described as an important factor in prognostic models (Templeton *et al.*, 1996; Croucher *et al.*, 1998). On the other hand, basal FSH concentration was strongly used by gynaecologists. This is in conflict with most of the literature on the subject, in which the predictive capacity of basal FSH in the prediction of pregnancy after IVF is reported to be limited (Scott *et al.,* 1989; Bancsi *et al.,* 2003; Abdalla and Thum 2004; Klinkert *et al.,* 2005). While FSH is related to egg quantity rather than quality, high basal FSH will reduce the number of eggs available an thus the chances of success.

The impact of the semen analysis in the decision to start treatment, i.e. IUI or IVF, was strong, whereas the relative contribution of semen analysis in the prediction of pregnancy after IUI was limited. Apparently, gynaecologists respond to poor semen motility with the decision for treatment, whereas the semen motility hardly influences their estimates of the chances of IUI outcome. In this study, only semen motility was varied, whereas other semen parameters were not assessed. Since other semen characteristics have predictive capacity for the prediction of IUI (Van Weert *et al.,* 2005), these factors could also play a role in the decisions of gynaecologists, but this was not assessed in the present study.

It is concluded that several of all routinely ordered tests in the fertility work-up contribute little, or not at all, to clinical decision-making according to existing medical evidence. Apparently, difficulties arise when clinical guidelines have to be implemented into clinical practice. Publishing studies and guidelines seem not to be sufficient to implement results of clinical research (Grol *et al.,* 2003). Obviously, behavioural changes require comprehensive approaches at various levels (doctor, team practice, hospital, wider environment), tailored to specific settings and target groups.

The main implication of the present study is that the process of implementation of evidence based guidelines as far as the fertility work-up is concerned, is not guaranteed with the correct ordering of tests by gynaecologists. Even if gynaecologists order the right tests, their interpretation may not be according to current knowledge. Solutions for this problem might be better training of gynaecologists or a more stringent use of information of clinical prediction rules, for example by integrating them in electronic decision support systems (Johnston *et al.,* 1994; Schiffman 2004; Garg *et al.,* 2005). It is planned to perform a randomized study in which couples are allocated by an electronic decision support system either to

management with treatment advise according to the existing prediction models, or to management according to the preference of the gynaecologists.

In conclusion, maternal age, duration of subfertility and FSH concentration seem to be the main factors in clinical decision-making of gynaecologists in subfertility. This study suggests that gynaecologists seem to overestimate the importance of basal FSH concentration and underestimate the importance of previous pregnancy in clinical decision-making as compared with their importance as reported in prediction models and guidelines. Education about the right use of test information in the fertility work-up is warranted, either by regular training or by electronic decision support systems.

Acknowledgements

The authors thank all participating gynaecologists for their contribution to this study.

References

Abdalla H and Thum MY. (2004) An elevated basal FSH reflects a quantitative rather than qualitative decline of the ovarian reserve. *Human Reproduction* 19,893-898.

Addelman S. (1962) Symmetrical and asymmetrical fractional factorial plans. *Technometrics* 4,47-58.

Akande VA, Hunt LP, Cahill DJ, Caul EO, Ford WC, and Jenkins JM. (2003) Tubal damage in infertile women: prediction using chlamydia serology. *Hum Reprod* 18,1841-1847.

Bancsi LF, Broekmans FJ, Mol BWJ, Habbema JDF, and Te Velde ER. (2003) 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* 79,1091-1100.

Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, and Te Velde ER. (2000) Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. *Fertil Steril* 73,552-557.

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Croucher CA, Lass A, Margara R, and Winston RM. (1998) Predictive value of the results of a first in-vitro fertilization cycle on the outcome of subsequent cycles. *Hum Reprod* 13,403-408.

Dutch Society of Obstetrics and Gynaecology. (2004) Guidline-Basic fertility work-up. *NVOG-richtlijn nr 1*,1-12.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, and Haynes RB. (2005) Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 293,1223-1238.

Grol R, Cluzeau FA, and Burgers JS. (2003) Clinical practice guidelines: towards better quality guidelines and increased international collaboration. *Br J Cancer* 89 Suppl 1,S4-S8.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Johnston ME, Langton KB, Haynes RB, and Mathieu A. (1994) Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med* 120,135-142.

Klinkert ER, Broekmans FJ, Looman CW, Habbema JDF, and Te Velde ER. (2005) The antral follicle count is a better marker than basal follicle-stimulating hormone for the selection of older patients with acceptable pregnancy prospects after in vitro fertilization. *Fertil Steril* 83,811-814.

Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Van der Veen F, and Bossuyt PMM. (1997) The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 67,1031-1037.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems. www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Schiffman M. (2004) Evidence-based screening and management guidelines address the realistic concerns of practicing clinicians and pathologists. *J Low Genit Tract Dis* 8,150-154.

Scott RT, Toner JP, Muasher SJ, Oehninger S, Robinson S, and Rosenwaks Z. (1989) Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. *Fertil Steril* 51,651-654.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Steures P, Van der Steeg JW, Mol BWJ, Eijkemans MJC, Van der Veen F, Habbema JDF, Hompes PGA, Bossuyt PMM, Verhoeve HR, Van Kasteren YM *et al*. (2004) Prediction of an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 82,45-51.

Templeton A, Morris JK, and Parslow W. (1996) Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 348,1402-1406.

Van Montfrans JM, Hoek A, Van Hooff MH, De Koning CH, Tonch N, and Lambalk CB. (2000) Predictive value of basal follicle-stimulating hormone concentrations in a general subfertility population. *Fertil Steril* 74,97-103.

Van Rooij IA, Broekmans FJ, Hunault CC, Scheffer G, Eijkemans MJC,DeJong FH, Themmen APN, and Te Velde ER. (2006) Use of ovarian reserve tests for the prediction of ongoing pregnancy in couples with unexplained or mild male infertility. *Reprod Biomed Online* 12,182-190.

Van Weert JM, Repping S, Van der Steeg JW, Steures P, Van der Veen F, and Mol BWJ. (2005) IUI in male subfertility: are we able to select the proper patients? *Reprod Biomed Online* 11,624-631.

HUMAN REPRODUCTION
2007; 22: 536 - 542

Nominated for the Promising Young Clinician Award

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Peter G.A. Hompes
Frank J. Broekmans
Thierry H.J.H.M. van Dessel
Patrick M.M. Bossuyt
Fulco van der Veen
Ben Willem J. Mol
for the CECERM investigators

chapter (05)

PREGNANCY IS PREDICTABLE

a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples

Abstract

Background Prediction models for spontaneous pregnancy may be useful tools to select subfertile couples that have good fertility prospects and should therefore be counselled for expectant management. We assessed the accuracy of a recently published prediction model for spontaneous pregnancy in a large prospective validation study.

Methods In 38 centres, we studied a consecutive cohort of subfertile couples, referred for an infertility work-up. Patients had a regular menstrual cycle, patent tubes and a total motile sperm count (TMC) > 3×10^6. After the infertility work-up had been completed, we used a prediction model to calculate the chance of a spontaneous ongoing pregnancy (www.freya.nl/probability.php).

The primary end-point was time until the occurrence of a spontaneous ongoing pregnancy within one year. The performance of the pregnancy prediction model was assessed with calibration, which is the comparison of predicted and observed ongoing pregnancy rates for groups of patients and discrimination.

Results We included 3021 couples of whom 543 (18%) had a spontaneous ongoing pregnancy, 57 (2%) a non-successful pregnancy, 1316 (44%) started treatment, 825 (27%) neither started treatment nor became pregnant, and 280 (9%) were lost to follow-up. Calibration of the prediction model was almost perfect. In the 977 couples (32%) with a calculated probability between 30% and 40%, the observed cumulative pregnancy rate at 12 months was 30%, and in 611 couples (20%) with a probability of ≥ 40%, this was 46%. The discriminative capacity was similar to the one in which the model was developed (c-statistic 0.59).

Conclusions As the chance of a spontaneous ongoing pregnancy among subfertile couples can be accurately calculated, this prediction model can be used as an essential tool for clinical decision-making and in counselling patients. The use of the prediction model may help to prevent unnecessary treatment.

Introduction

Infertility or involuntary childlessness is one of life's great catastrophes, which happens to approximately one out of ten couples (Gibbons 1911; Hull *et al.,* 1985; Wang *et al.,* 2003; Gnoth *et al.,* 2003). Most of these couples seek medical help (Evers 2002). Assessment of the fertility potential in an infertility work-up is then the first step (National institute for clinical excellence 2004). After the completion of the infertility work-up, it is essential to distinguish subfertile couples in whom prognosis of spontaneous pregnancy is poor and fertility treatment is mandatory from subfertile couples who still have a good prognosis to conceive spontaneously (Te Velde and Cohlen 1999). Prediction models for spontaneous pregnancy may be useful tools to achieve this (Wasson *et al.,* 1985).

In the last decade, three prediction models for the prediction of spontaneous pregnancy in subfertile couples have been published (Eimers *et al.,* 1994b; Collins *et al.,* 1995; Snick *et al.,* 1997). In 1994, Eimers *et al.* developed a model that was based on more than 900 subfertile couples who visited a university fertility centre in the Netherlands before the era of IVF. In 1995, Collins *et al.* developed a model that was based on tertiary care subfertile couples who visited one of the eleven participating IVF centres in Canada. Finally, in 1997, Snick *et al.* developed a model that was based on data derived from a secondary care fertility centre, on the Walcheren peninsula in the Netherlands.

To broaden the empirical basis of these pregnancy predictions models, the data of the three prediction models were pooled and integrated in a synthesis model (Hunault *et al.,* 2004). This synthesis model is a pregnancy prediction model that contains the variables female age, duration of subfertility, subfertility being primary or secondary, semen motility and referral status, and is available in a version with and without the post-coital test as predictor.

At the moment, none of the above-described pregnancy prediction models have been implemented in fertility guidelines or overviews (National institute for clinical excellence 2004; Boston (MA) 2003; Bloomington (MN): Institute for Clinical Systems Improvement (ICSI) 2004; Bagshawe and Taylor 2003), even though it has been argued that couples must be informed about their chances of conceiving spontaneously before considering any treatment (Te Velde and Cohlen 1999). The question is, therefore, why these prediction models have not yet been recommended in guidelines. One important reason may be the fact that the general

applicability of these pregnancy prediction models has not been assessed in prospective validation studies in a wide variety of settings (Wasson *et al.,* 1985). This can be done by external validation, an essential and final step in the development of prediction models (Wasson *et al.,* 1985; Bleeker *et al.,* 2003), because prediction models tend to be overoptimistic when applied in other populations than the one in which they were developed (Stolwijk *et al.,* 1996; Stolwijk *et al.,* 1998).

The aim of this study was to validate the synthesis prediction model for spontaneous pregnancy in prospectively collected data from an external population. We also assessed the ability of the pregnancy prediction model to differentiate couples with a good chance to conceive from couples in whom treatment should be started because their prospects are too poor.

Materials and methods

The study was designed as a prospective cohort study performed in 38 hospitals in the Netherlands. The local ethics committee of each participating centre gave Institutional Review Board approval for this study. Between January 2002 and February 2004, we included consecutive subfertile couples who had not been evaluated previously for subfertility. All couples were referred by their general practitioner and underwent an infertility work-up, consisting of a fertility history, semen analysis, PCT, assessment of ovulation and assessment of the Fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (Dutch Society of Obstetrics and Gynaecology 2004). Couples referred by a gynaecologist were not included in this study, because these couples were referred for fertility treatment rather than for evaluating their fertility potential.

The duration of subfertility was defined as the period between the time the couple had an active child wish and the moment at which the infertility work-up was completed. If the couple had a previous pregnancy that had not resulted in a live birth, the duration of subfertility was defined as the period between the renewed active child wish after this pregnancy and the moment at which the infertility work-up was completed. Female age was calculated at the time the infertility work-up ended. At the time of the study no treatment was offered to women older than 40 years. Subfertility was considered to be secondary if a woman had conceived in the current or in a prior partnership, regardless of the pregnancy outcome.

The menstrual cycle was considered regular if the duration of the cycle was between 23 and 35 days, with an intercycle variation of less than 8 days. Presence of ovulation was confirmed by means of a basal body temperature (BBT) chart, a mid-luteal serum progesterone, or by sonographic monitoring of the cycle. Women with an ovulation disorder, i.e. an irregular menstrual cycle or anovulation, were not included in the study.

Semen analysis was performed at least once. Couples in whom the man had a total motile sperm count (TMC) $< 3 \times 10^6$ were excluded from the analysis.

At least one PCT was performed during the infertility work-up (Eimers *et al.,* 1994a; Snick *et al.,* 1997; Glazener *et al.,* 2000; Van der Steeg *et al.,* 2004). The PCT could be planned based on the BBT and cycle length or on repetitive ultrasound findings. The PCT was judged to be normal if at least one progressively motile spermatozoon was seen in one of five high-power

fields at x400 magnification. All other PCT results were considered to be abnormal.

Tubal pathology was assessed by a Chlamydia Antibody Test (CAT), or directly assessed by hysterosalpingography (HSG) or laparoscopy. In case of a positive CAT, the tubal status was subsequently evaluated with HSG or laparoscopy, whereas in cases with a negative CAT, tubal pathology was considered absent (Mol *et al.,* 1997). The CAT could be tested with immune fluorescence technique (IF) or with enzyme immune assays (EIA) (BioMerieu, Paris, France; Medac GmbH, Wedel, Germany; Savyon Diagnostics, France). The CAT was considered to be positive for IF if the titre was > 1:16 and for ELISA if the level was > 1.1. Women with one- or two-sided tubal pathology were excluded from the further follow-up and analysis.

After the completion of the infertility work-up a probability of spontaneous pregnancy within one year was calculated. To do so, we used the pregnancy prediction model developed by Hunault *et al. (*Hunault *et al.,* 2004). This model is based on a Cox regression model developed to predict spontaneous pregnancy within one year, leading to live birth, fitted to the data used in the construction of previous prediction models by Eimers, Collins and Snick (Eimers *et al.,* 1994b; Collins *et al.,* 1995; Snick *et al.,* 1997), and as such is called the 'synthesis' model. It includes six prognostic variables: female age, duration of subfertility, female subfertility being primary or secondary, percentage motile spermatozoa of the first semen analysis, result of the PCT and referral status (being referred by general practitioner or gynaecologist). Each variable is converted in a point score. The total point score of each couple corresponds to a prognosis of a spontaneous ongoing pregnancy. The computer model can be used with following URL: www.freya.nl/probability.php. The synthesis prediction model is available in two variants; one includes the PCT, whereas the other does not include the PCT. In this manuscript, we evaluate the model without the PCT, which we refer to as the prediction model for spontaneous pregnancy. The results of the model with the PCT are presented in the discussion.

Couples in whom the prognosis for spontaneous pregnancy was ≥ 40% within 12 months were counselled for expectant management for a period of at least six months. After six months of expectant management, it was up to the couples to decide whether to start treatment or to wait for a longer period. Couples with a prognosis below 40% were counselled for treatment according to the national fertility guidelines (Dutch Society of Obstetrics and Gynaecology 2000; Dutch Society of Obstetrics and Gynaecology 1998).

Follow-up started at the completion of the infertility work-up and ended after 12 months. Primary end-point in this study was time to spontaneous conception resulting in an ongoing pregnancy. The first day of that menstrual cycle was considered to mark the end of time until spontaneous conception. Spontaneous ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal sonography at a gestational age of at least 12 weeks, resulting from a treatment-independent conception. Time to pregnancy was censored at the moment treatment had been started within 12 months after counselling or at the last date of contact during follow-up, when the couple had no ongoing pregnancy. In case of a miscarriage or ectopic pregnancy, follow-up continued. The time to pregnancy in these couples was not censored at the time of the unsuccessful pregnancy. For all couples who were lost to follow-up, the general practitioner was sent a questionnaire and asked about the fertility status of the couple.

Analysis

Missing data of the predictive variables were imputed ('filled in'), because deleting them would lead to a loss of statistical power in multivariable analysis and -more seriously- potentially biased results (Little and Rublin 1987; Schafer 1997). We generated an imputed dataset, using ' 'aRegImpute' imputation function in Splus 6.0. This is an efficient implementation of Bayesian multiple imputation, a recommended state of the art method (Schafer and Graham 2002). The predictors were female age, duration of subfertility, previous pregnancy, semen motility, PCT and the referral status. We generated an imputed dataset, using the 'aRegImpute' imputation function in S-plus® 6.0.

We evaluated the calibration of the prediction model for spontaneous pregnancy (Hunault *et al.,* 2004). Calibration is the degree of comparison between observed and predicted event rates for groups of patients (Altman and Royston 2000). Calibration was assessed by comparing the mean predicted probability with the mean observed fraction of ongoing pregnancies at 12 months, in 10 subgroups. For this purpose, the cohort was split into 10 groups based on the deciles of the calculated probabilities. Per group, the mean predicted probability as well as the mean observed fraction, calculated with the Kaplan Meier method, was calculated. For all groups, we plotted the predicted and observed means in a calibration plot. We calculated the slope and its confidence interval (CI) for the regression curve of the plot. In case of a perfect calibration, all predictions and observations would be located on the line of equality ($x = y$), and the slope

would be unity (slope = 1) with a *p*-value > 0.05 (not significantly different from 1).

Discrimination is a measurement of the ability to distinguish between patients who do and do not experience the event of interest, in this case getting pregnant without treatment (Altman and Royston 2000). Discriminative capacity was assessed with receiver operation characteristic (ROC) analysis. We also calculated the *c*-statistic. This statistic is comparable to the area under the ROC-curves (AUC) but corrects for the fact that some couples were censored before follow-up of twelve months (Harrell, Jr. *et al.,* 1996).

Calculations were performed with SPSS® 12.0 (SPSS Inc., Chicago, IL) and S-plus® 6.0 (MathSoft Inc., Seattle, WA) programs.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

In the period between 1 January 2002 and 1 February 2004 we registered 5591 subfertile couples. Imputation was done on all patients who had at most two missing values in the six core prognosticators for spontaneous pregnancy. In total, 4.3% data points were missing and subsequently imputed. Of these 5591 patients, 948 had a severe male factor, 311 had two-sided tubal pathology and 1311 had one-sided tubal pathology, leaving 3021 couples for inclusion. Baseline characteristics of these couples are summarized in table 1.

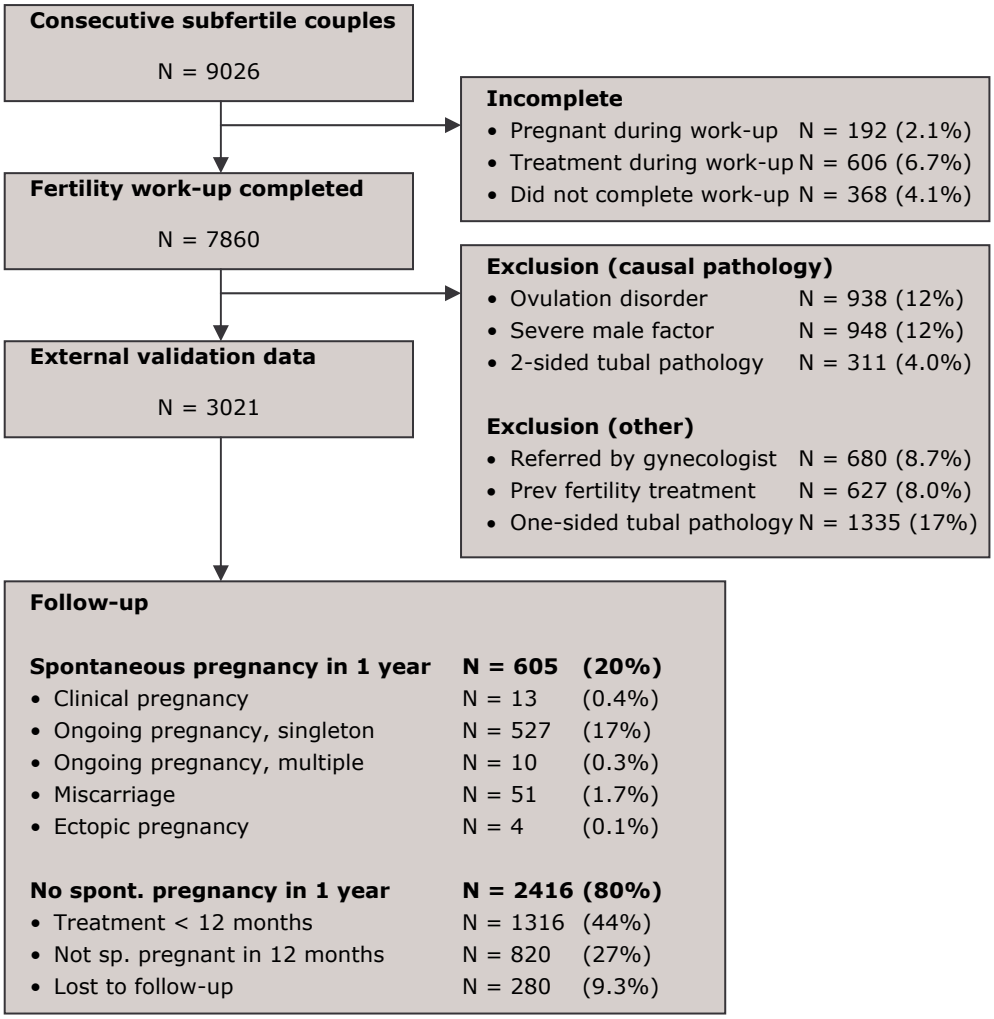
Table 1 Baseline characteristics of the 3021 included subfertile couples.

	Mean	**Median**	**5th-95th percentile**
Female age (years)	32.2		25–39
Male age (years)	34.8		27–44
Duration of subfertility (years)		1.7	1.0–3.9
Subfertility, primary (*n*) / (%)	2013		67%
Semen analysis – TMC (10^6)		55.2	4.0–315
PCT, normal (*n*) / (%)	2023		67%
Cycle length (days)		28.1	24–33
FSH (U/L)		6.4	3.5–12
BMI (kg/m^2)		23.0	19–34

BMI = body mass index; FSH = follicular stimulating hormone; PCT = postcoital test; TMC = total motile sperm count.

In 1917 couples (64%) the estimated probability of a spontaneous pregnancy was < 40% and possible treatment was discussed with the couples. In this group, 17% became pregnant without treatment. In 1104 couples (36%), the probability of a spontaneous pregnancy was 40% or higher and expectant management for at least six months was recommended. In this group, 52% of the couples fulfilled the six-month period of expectant management, whereas 30% were treated before six months.

Figure 1 Study profile.



The follow-up status of all patients at 12 months is shown in figure 1. We completed follow-up for 2741 couples (91%). Of the 3021 couples, 543 (18%) had a spontaneous ongoing pregnancy within one year, including ten multiple pregnancies (1.8%). Unsuccessful pregnancies occurred in 57 couples (1.9%), of which 53 resulted in a miscarriage, and four resulted in an ectopic pregnancy, i.e. 9.5% and 0.7% of all pregnancies, respectively.

The follow-up status of all patients at 12 months is shown in figure 1. We completed follow-up for 2741 couples (91%). Of the 3021 couples, 543 (18%) had a spontaneous ongoing pregnancy within one year, including ten multiple pregnancies (1.8%). Unsuccessful pregnancies occurred in 57 couples (1.9%), of which 53 resulted in a miscarriage, and four resulted in an ectopic pregnancy, i.e. 9.5% and 0.7% of all pregnancies, respectively. Within 12 months 1316 (44%) of all couples started treatment, whereas 825 (27%) did neither start treatment yet nor became pregnant within the follow-up period.

The mean probability of a spontaneous pregnancy as calculated with the prediction model was 0.32 (5^{th} and 95^{th} percentiles: 0.16 and 0.55). The overall cumulative fraction of untreated couples with a spontaneous pregnancy at 12 months was 29.5% (95% CI: 27% – 32%) (figure 2).

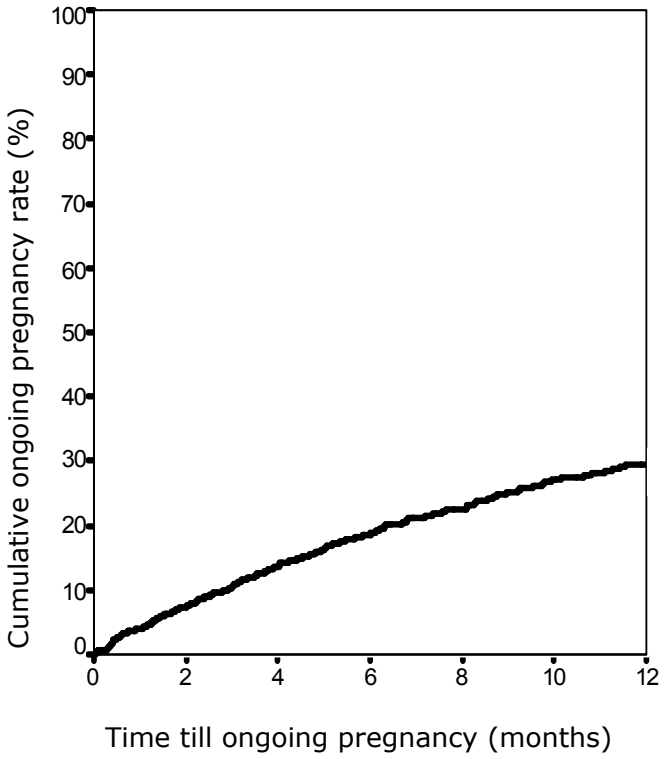


Figure 2 Results of the Kaplan Meier analysis.

Results of the calibration analysis of the prediction model for spontaneous pregnancy are shown in figure 3. This figure shows the association between the mean calculated probability of spontaneous ongoing pregnancy and the mean observed fraction of spontaneous ongoing pregnancies for each of the 10 decile groups. The prediction model for spontaneous pregnancy had a good calibration, with a slope of 0.82 (95% CI 0.6 – 1.0, *p*-value 0.08). In the 977 couples (32%) with a calculated probability, between 30% and 40%, the observed cumulative pregnancy rate at 12 months was 30%, and in 611 couples (20%) with a probability of ≥ 40%, this was 46%.

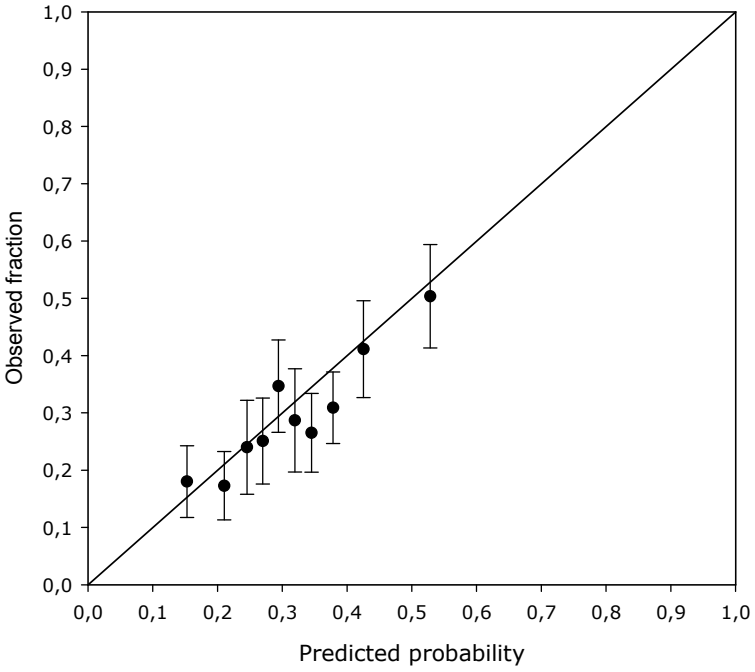


Figure 3 Calibration plot of the prediction model for spontaneous pregnancy.

For an assessment of the discriminative capacity, we calculated a *c*-statistic of 0.59 for the prediction model for spontaneous pregnancy, both for the version with and without the PCT. The discriminative capacity was similar to the one in which the model was developed, where the *c*-statistic had values of 0.63 and 0.59 for the version with and without the PCT, respectively.

Discussion

Our study was designed to validate the recently developed synthesis prediction model for spontaneous pregnancy in subfertile couples in routine clinical practice. For a period of two years, 38 hospitals used this prediction model to counsel subfertile couples for expectant or active management according to their predicted probabilities. We showed that the prediction model for spontaneous pregnancy is able to provide accurate probabilities of ongoing pregnancy over the whole range of subfertile couples.

We feel that this study is important for three reasons. First, the final step in the development of a prediction model for spontaneous pregnancy, i.e. external validation, has been performed. This essential step is often refrained from the development of prediction models, because prospective validation studies are costly and time-consuming (Wasson *et al.,* 1985). Second, this study shows that a prediction model for spontaneous pregnancy can be applied in a general subfertile population, without loosing its accuracy, regardless of the setting of the hospital. Third, the information produced by the prediction model for spontaneous pregnancy is accurate. The accurate predictions that we report here justify its use in counselling individual patients. Such an accurate counselling can improve patient care because it enables both doctors and patients to whether to start fertility treatment. For example, a > 40% chance of spontaneous pregnancy in one year may justify expectant management, because it is not to be expected that assisted reproduction will increase the already high pregnancy rates in this group. Of course, one should also consider that an expectant management is followed by treatment in case pregnancy does not occur. Pros and cons can be discussed between doctor and patient, and a shared and evidence based decision can be made.

Originally, the prediction model was developed to predict spontaneous pregnancy leading to live birth. In this study, we validated the model, using ongoing pregnancy as the primary end-point, because at the time of the analysis of this study follow-up of pregnancies until live birth was not complete for all couples. Because less than 5% of ongoing pregnancies do not lead to live birth (Yudkin *et al.,* 1987; Feldman 1992; Hilder *et al.,* 1998), we assume that the prediction model is also well calibrated for the prediction of spontaneous pregnancy leading to live birth.

The performance of the prediction model was assessed by evaluating calibration and discrimination. However, we want to stress that in the assessment of performance of a prediction model, calibration is more

important than discriminative capacity. The latter is more appropriate for evaluating diagnostic tests (Mol *et al.,* 2005). In subfertile couples, it is more important to accurately estimate whether a couple has a high or low chance to conceive than to know exactly which couple will conceive and which not (Cook *et al.,* 2006).

This study was performed in a multi-centre setting to assess whether the prediction model for spontaneous pregnancy is applicable to a general subfertile population. A consequence of the multi-centre design was that there was heterogeneity in the performance of tubal assessment. Some clinics completed the fertility work-up by assessing the tubal status with a CAT (Mol *et al.,* 1997; Akande *et al.,* 2003), whereas other clinics always performed either a HSG or a laparoscopy as final part of the fertility work-up. Therefore, the duration of the fertility work-up in the latter clinics was prolonged, influencing both female age and duration of subfertility. Stratified analyses for these two types of clinics showed comparable baseline characteristics and virtual similar cumulative ongoing pregnancy rates after one year (29% and 30%, respectively, data not further shown). We therefore combined these two types of clinics in our analyses.

In this study, we focussed on subfertile couples being referred by their general practitioner, and excluded couples referred to third care clinics. We assumed that the latter group was referred for fertility treatment and had an assisted reproductive technique indication. Therefore, the use of a prediction model would not have consequences in this group. As a result, there was no variance in the variable 'referral status', and subsequently, it only contributed to the baseline probability for all couples.

We also validated the original models of Eimers, Collins and Snick on our data as well as the alternative prediction model that includes the PCT as predictor (Eimers *et al.,* 1994b; Collins *et al.,* 1995; Snick *et al.,* 1997). The model of Eimers was shown to underestimate the pregnancy rates in couples with poor prognoses (0-20%) and overestimate them in couples with good prognoses (≥ 40%). The calibration was poor. The model of Collins underestimated the pregnancy rates in all couples regardless of their prognoses. The calibration was poor. The model of Snick also showed a poor calibration in a validation study. The prediction model for spontaneous pregnancy, including the PCT, tended to underestimate the chance of conception in couples with a poor prognosis (between 10% and 25%), but overestimated those chances in couples with a good prognosis. Overall calibration was moderate with a slope of 0.58 (95% CI 0.4 – 0.7, *p*-value < 0.01). All the three original models and the alternative synthesis model

with the PCT performed less than the prediction model validated in this study.

Since the introduction of IVF in 1978, its use has increased tremendously. In the USA alone, more than 58000 IVF cycles are performed annually (ASRM 2002). The most important drawback of this phenomenon is high multiple pregnancy rates (Jain *et al.,* 2002; ASRM 2004), with concomitant complications for both mother and child (Basso and Olsen 2005; Mitchell 2002). The reduction of the number of transferred embryos has been proposed as a key towards the multiple pregnancy problem (Bhattacharya and Templeton 2000; Guzick 2002). Furthermore, IVF is costly. Assuming an IVF cycle to cost $10 000, the average direct cost per baby after IVF is approximately $40 000 (Guzick 2002; Jain *et al.,* 2002). The use of prognostic models with tailored expectant management in couples who are very likely to conceive without assisted reproduction is in our opinion a powerful tool to limit the use of IVF to those subfertile couples with a low likelihood to conceive without IVF.

At this moment, neither clinical guidelines nor current overviews from the American College of Obstetrics and Gynaecology (ACOG) or the Royal College of Obstetrics and Gynaecology (RCOG) (2004) mention the possibility of assessment of the prognosis in the fertility work-up and subsequent counselling according to this prognosis (National institute for clinical excellence 2004; Bagshawe and Taylor 2003). On the basis of the results from present study, we believe that the use of prediction models should be discussed in current fertility guidelines.

In clinical practice, the pregnancy prediction model can be used in two ways. The clinician can use a paper score chart (table 2 and figure 4) or a computer version of the prediction model, which calculates the chance to conceive automatically, once the predictive parameters have been entered. This model is now available on the Internet (www.freya.nl/probability.php). This format enables gynaecologists all over the world to predict chances on spontaneous pregnancy in a busy practice within a few seconds.

In conclusion, the recently developed synthesis prediction model for spontaneous pregnancy predicts accurately and performs well in a general population. The use of this model in the interpretation of the fertility work-up enables one to identify couples who have a good chance to conceive without treatment. The use of this prediction model may prevent unnecessary treatment. We suggest use of this pregnancy prediction model as an essential tool in counselling subfertile couples.

CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Investigators, participating centres in The Netherlands:
Y.M. van Kasteren; Alkmaar, Medisch Centrum Alkmaar; P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis; W.A. Schöls; Amersfoort, Meander Medisch Centrum; M.H. Mochtar; Amsterdam, Academisch Medisch Centrum; G.L.M. Lips; Amsterdam, BovenIJ Ziekenhuis; J. Dawson; Amsterdam, St. Lucas Adreas Ziekenhuis; H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis; S. Milosavljevic; Amsterdam, Slotervaart Ziekenhuis; P.G.A. Hompes; Amsterdam, Vrij Universiteit Medisch Centrum; L.J. van Dam; Apeldoorn, Gelre Ziekenhuis; A.V. Sluijmer; Assen, Wilhelmina Ziekenhuis; H.E. Bobeck; Beverwijk, Rode Kruis Ziekenhuis; R.E. Bernardus; Blaricum, Ziekenhuis Gooi-Noord; M.C.S. Vermeer; Breda, Amphia Ziekenhuis Breda; J.P. Dörr; Den Haag, Medisch Centrum Haaglanden, locatie Westeinde; P.J.Q. van der Linden; Deventer, Ziekenhuis Deventer; H.J.M. Roelofs; Dordrecht, A. Schweitzer Ziekenhuis; J.M. Burggraaff; Emmen, Scheper Ziekenhuis Emmen; G.J.E. Oosterhuis; Enschede, Medisch Spectrum Twente; M.H. Schouwink; Geldrop, St. Anna Ziekenhuis; P.X.J.M. Bouckaert; Heerlen, Atrium Medisch Centrum; F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis; C.J.C.M. Hamilton; 's Hertogenbosch, Jeroen Bosch Ziekenhuis; M. van Hoven; Hilversum, Tergooiziekenhuizen; M.H. Emanuel; Hoofddorp, Spaarne Ziekenhuis; C.N.M. Renckens; Hoorn, Westfries Gasthuis; J.A. Land; Maastricht, Academisch Ziekenhuis Maastricht; J.H. Schagen-Van Leeuwen; Nieuwegein, St. Antonius Ziekenhuis; J.A.M. Kremer; Nijmegen, UMC St. Radboud; C. van Katwijk; Purmerend, Waterland Ziekenhuis; M.H.A. van Hooff; Rotterdam, St. Franciscus Gasthuis; H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis; F.J.M. Broekmans; Utrecht, UMC Utrecht; H.J.L.A. Ruis; Veghel, Ziekenhuis Bernhoven; C.A.M. Koks; Veldhoven, Máxima Medisch Centrum; P. Bourdrez; Venlo/Venray, Vie Curi Medisch Centrum; W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum; B.J. Cohlen; Zwolle, Isala Klinieken.

Appendix

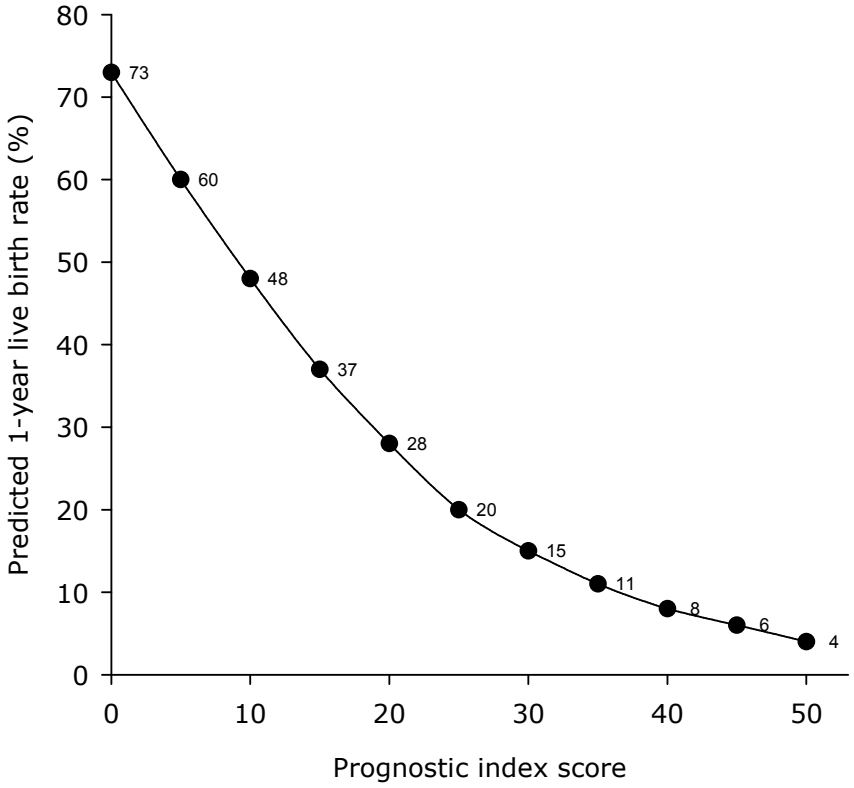


Figure 4 Prognostic score chart (Hunault *et al.* 2004).

Example: a couple with a 32-year-old woman, with primary subfertility of 1 year duration, with 20% progressive motile sperm, referred by the general practitioner has a prognostic index score of: 6 + 0 + 0 + 4 + 0 = 10, corresponding with a predicted 1-year live birth rate of 48%. Computer model can by used with following URL: www.freya.nl/probability.php.

Table 2 Prediction model for spontaneous pregnancy within 12 months, leading to live birth (Hunault *et al.* 2004).

Variable	Regression coefficient (β)	Categorical	Score chart index
Intercept	0.17		
Maternal age if ≤ 31 years (per year older) (AGE 1)	-0.03 (β1)	21-25	0
		26-31	2
Maternal age if > 31 years (per year older than 31) (AGE 2)	-0.08 (β2)	32-35	6
		36-37	9
		38-39	11
		40-41	12
Duration of subfertility (per year longer) (DUR)	-0.19 (β3)	1	0
		2	2
		3-4	5
		5-6	9
		7-8	13
Primary subfertility of the couples (primary = 1, secondary = 0) (PRIM)	-0.58 (β4)	Primary	0
		Secondary	6
Progressive motile semen (%) (PROG)	0.008 (β5)	≥ 60	0
		40-59	2
		20-39	4
		0-19	6
Referral status (tertiary couple = 1, secondary couple = 0) (REF)	-0.25 (β6)	Secondary	0
		Tertiary-care	4
		Sum	

$P_{pregnancy} = 1\text{-}0.17 \times e^{((AGE\ 1\ X\ \beta1) + (AGE\ 2\ X\ \beta2) + (DUR\ x\ \beta3) + (PRIM\ X\ \beta4) + (PROG\ X\ \beta5) + (REF\ X\ \beta6))}$

References

Akande VA, Hunt LP, Cahill DJ, Caul EO, Ford WC, and Jenkins JM. (2003) Tubal damage in infertile women: prediction using chlamydia serology. *Hum Reprod* 18,1841-1847.

Altman DG and Royston P. (2000) What do we mean by validating a prognostic model? *Stat Med* 19,453-473.

ASRM. (2002) Assisted reproductive technology in the United States: 1998 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. *Fertil Steril* 77,18-31.

ASRM. (2004) Multiple pregnancy associated with infertility therapy. *Fertil Steril* 82,153-157.

Bagshawe A and Taylor A. (2003) ABC of subfertility. Counselling. *BMJ* 327,1038-1040.

Basso O and Olsen J. (2005) Subfecundity and neonatal mortality: longitudinal study within the Danish national birth cohort. *BMJ* 330,393-394.

Bhattacharya S and Templeton A. (2000) In treating infertility, are multiple pregnancies unavoidable? *N Engl J Med* 343,58-60.

Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, and Moons KG. (2003) External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 56,826-832.

Bloomington (MN): Institute for Clinical Systems Improvement (ICSI). (2004) Diagnosis and management of basic infertility. www.guideline.gov.

Boston (MA). (2003) Infertility. A guide to evaluation, treatment, and counseling. www.guideline.gov.

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Cook NR, Buring JE, and Ridker PM. (2006) The Effect of Including C-Reactive Protein in Cardiovascular Risk Prediction Models for Women. *Ann Intern Med* 145,21-29.

Dutch Society of Obstetrics and Gynaecology. (1998) Guideline-Indications for invitro fertilization (IVF). *NVOG-richlijn nr 9,*1-7.

Dutch Society of Obstetrics and Gynaecology. (2000) Guideline-Intra-uterine insemination (IUI). *NVOG-richtlijn nr 29,*1-8.

Dutch Society of Obstetrics and Gynaecology. (2004) Guideline-Basic fertility work-up. *NVOG-richtlijn nr 1,*1-12.

Eimers JM, Te Velde ER, Gerritse R, Van Kooy RJ, Kremer J, and Habbema JDF. (1994a) The validity of the postcoital test for estimating the probability of conceiving. *Am J Obstet Gynecol* 171,65-70.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994b) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Evers JL. (2002) Female subfertility. *Lancet* 360,151-159.

Feldman GB. (1992) Prospective risk of stillbirth. *Obstet Gynecol* 79,547-553.

Gibbons RA. (1911) A lecture on sterility, its etiology and treatment. *London: J&A Churchill*.

Glazener CM, Ford WC, and Hull MG. (2000) The prognostic power of the post-coital test for natural conception depends on duration of infertility. *Hum Reprod* 15,1953-1957.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, and Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18,1959-1966.

Guzick DS. (2002) Should insurance coverage for in vitro fertilization be mandated? *N Engl J Med* 347,686-688.

Harrell FE, Jr., Lee KL, and Mark DB. (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15,361-387.

Hilder L, Costeloe K, and Thilaganathan B. (1998) Prolonged pregnancy: evaluating gestation-specific risks of fetal and infant mortality. *Br J Obstet Gynaecol* 105,169-173.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, and Desai KM. (1985) Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 291,1693-1697.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Jain T, Harlow BL, and Hornstein MD. (2002) Insurance coverage and outcomes of in vitro fertilization. *N Engl J Med* 347,661-666.

Little,R.J.A. and Rublin,D.B. (1987) Statistical analysis with missing data. Wiley, New York.

Mitchell AA. (2002) Infertility treatment-more risks and challenges. *N Engl J Med* 346,769-770.

Mol BWJ, Coppus S.F., Van der Veen F, and Bossuyt PMM. Evaluating Predictors for the Outcome of Assisted Reproductive Technology: ROC-Curves Are Misleading; Calibration is not! Fertility and Sterility 84, s253-s254. 2005.

Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Van der Veen F, and Bossuyt PMM. (1997) The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 67,1031-1037.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems.
www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Schafer JL (1997) Analysis of incomplete multivariate data. Chapman & Hall, London.

Schafer JL and Graham JW. (2002) Missing data: our view of the state of the art. *Psychol Methods* 7,147-177.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Stolwijk AM, Straatman H, Zielhuis GA, Jansen CA, Braat DD, Van Dop PA, and Verbeek AL. (1998) External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. *Hum Reprod* 13,3542-3549.

Stolwijk AM, Zielhuis GA, Hamilton CJ, Straatman H, Hollanders JM, Goverde HJ, Van Dop PA, and Verbeek AL. (1996) Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. *Hum Reprod* 11,2298-2303.

Te Velde ER and Cohlen BJ. (1999) The management of infertility. *N Engl J Med* 340,224-226.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Van der Veen F, Bossuyt PMM, Hompes PGA, and Mol BWJ. (2004) Should the post-coital test (PCT) be part of the routine fertility work-up? *Hum Reprod* 19,1373-1379.

Wang X, Chen C, Wang L, Chen D, Guang W, and French J. (2003) Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79,577-584.

Wasson JH, Sox HC, Neff RK, and Goldman L. (1985) Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 313,793-799.

Yudkin PL, Wood L, and Redman CW. (1987) Risk of unexplained stillbirth at different gestational ages. *Lancet* 1,1192-1194.

FERTILITY AND STERILITY
2007; [Epub ahead of print]

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Patrick M.M. Bossuyt
Peter G.A. Hompes
Fulco van der Veen
Ben Willem J. Mol
for the CECERM investigators

chapter (06)

PREDICTIVE VALUE OF PREGNANCY HISTORY IN SUBFERTILE COUPLES

results from a nationwide cohort study in the Netherlands

Abstract

Objective To assess whether pregnancy history can predict the occurrence of spontaneous ongoing pregnancy in subfertile couples.
Design Prospective cohort study.
Setting Thirty fertility centers in the Netherlands.
Patients Subfertile, ovulatory women with at least one patent tube and male partners without severely impaired semen quality.
Interventions Fertility work-up, including a detailed pregnancy history.
Main outcome measures Spontaneous ongoing pregnancy.
Results We included 4445 couples, of whom 793 (18%) had a spontaneous ongoing pregnancy within one year of follow-up. Previous live birth and miscarriage in current partnership were both associated with higher fecundity as compared with primary infertility (Hazard rate ratios for spontaneous pregnancy (HR): 1.4 [95% CI 1.2 – 1.7] and 1.3 [95% CI 1.0 – 1.5], respectively). Pregnancies in a woman's previous partnerships did not affect the fecundity of the couple. A pregnancy in previous partnership of the male partner was associated with lower fecundity (HR 0.76 [95% CI 0.58 – 0.99]). A previous pregnancy after fertility treatment also was associated with lower fecundity (HR 0.52 [95% CI 0.30 – 0.90]).
Conclusions Accurate prediction of future fertility of a couple requires an exact assessment of the fertility history of both partners.

Introduction

Approximately 10% of the couples who want to have a child fail to conceive within one year of regular, unprotected intercourse (Hull *et al.,* 1985; Wang *et al.,* 2003; Gnoth *et al.,* 2003; Gnoth *et al.,* 2005). Of these couples, 15 to 30% have conceived before with their current partner, and 5 to 10% of women conceived in previous partnerships (Eimers *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997).

At present, a previous pregnancy is considered to be a predictor for the occurrence of a new spontaneous pregnancy in subfertile couples. In the existing prediction models for spontaneous pregnancy, a previous pregnancy is expected to increase the probability of pregnancy with a factor 1.5 to 2 as compared with primary subfertility (Eimers *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997).

Unfortunately, the studies on which these models are based do not report on the characteristics of previous pregnancies with respect to the localization of the pregnancy (intrauterine or not), the outcome (live birth or not), whether they occurred in the current partnership or in previous partnerships, and whether the pregnancy occurred spontaneously or after assisted reproductive treatment (ART). This is worrisome because it is likely that these factors do have an impact on the subfertile couple's future fecundity. For example, it is unlikely that a previous ectopic pregnancy has the same predictive impact as a previous live birth, miscarriage or induced abortion as an ectopic pregnancy harms the fallopian tube, which may reduce future fertility.

We therefore assessed in a large prospective cohort study whether the outcome and origin of a prior pregnancy was related to the probability of spontaneous ongoing pregnancy in subfertile couples.

Materials and methods

The study was designed as a prospective multicenter cohort study performed in 30 hospitals in the Netherlands. The local ethics committee of each participating center gave institutional review board approval for this study, and the authors had no conflict of interest. We included consecutive subfertile couples between January 2002 and February 2004. All couples underwent a basic fertility work-up, consisting of a detailed fertility history, semen analysis, assessment of ovulation, postcoital test (PCT), and assessment of the fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and Gynecology (Dutch Society of Obstetrics and Gynaecology 2004).

With each couple, both the man and the woman were asked about any previous pregnancies that had occurred in current or previous partnerships. We recorded whether previous pregnancies were live births, miscarriages, induced abortions, or ectopic pregnancies. The couples were also asked whether the previous pregnancies had been achieved spontaneously or after fertility treatment, and whether the previous pregnancies had occurred in current or a previous partnership.

We distinguished between previous live birth, previous miscarriage, previous induced abortion, or previous ectopic pregnancy. A previous live birth was defined as a child still living 1 week after birth. A previous miscarriage was defined as an intrauterine pregnancy that resulted in a pregnancy failure before 16 weeks of gestation. A previous induced abortion was defined as an intrauterine pregnancy, which the woman had decided for abortion. An ectopic pregnancy was defined as a pregnancy located outside the uterine cavity that had been treated surgically or with methotrexate.

Fertility treatment was defined as the use of ovulation induction, intrauterine insemination with or without ovarian hyperstimulation, or treatment with in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI).

Duration of subfertility was defined as the period between the time the couple had an active desire to conceive and the moment at which the infertility work-up was completed. Female age was calculated at the time the infertility work-up ended. At the time of the study, no treatment was offered to women older than 40 years. Referral status was distinguished as referred by the general practitioner or referred by another gynecologist.

The menstrual cycle was considered regular if the duration of the cycle was between 23 and 35 days, with an intercycle variation of less than 8 days.Ovulation was assessed by means of a basal body temperature chart, a midluteal serum progesterone, or by sonographic monitoring of the cycle. Women with an irregular menstrual cycle or anovulation were excluded.

Semen analysis was performed at least once. The prewash total motile sperm count (TMC) was calculated from semen volume, semen concentration and percentage progressive motility. Men with a TMC < 3 x 10^6 were excluded from the study.

Tubal pathology was assessed by a chlamydia antibody test (CAT) or directly assessed by hysterosalpingography (HSG) or laparoscopy. In case of a positive CAT the tubal status was subsequently evaluated with HSG or laparoscopy; in cases of a negative CAT result, tubal pathology was considered absent (Mol *et al.,* 1997). The CAT result could be verified with an immune fluorescence technique (IF) or with enzyme immune assays (EIA) (BioMerieu, Paris, France; Medac GmbH, Wedel, Germany; Savyon Diagnostics Ltd., France). The CAT was considered positive by immune florescence if the titer was > 1:16 and for enzyme-linked immunosorbent assay (ELISA) if the level was > 1.1. Women with two-sided tubal pathology were excluded from the study.

After completion of the fertility work-up, the chance of spontaneous pregnancy within one year leading to live birth was calculated with a validated prediction model (www.freya.nl/probability.php) (Hunault *et al.,* 2004; Van der Steeg *et al.,* 2007). Couples in whom the probability of spontaneous pregnancy was ≥ 40% within 12 months were counseled for expectant management for a period of at least six months. After six months of expectant management, it was up to the couple to decide to start treatment or to wait for a longer period. Couples with a probability below 30% were counseled for treatment according to the national fertility guidelines, and those with a probability between 30% and 40% were asked to participate in a randomized clinical trial comparing intrauterine insemination (IUI) and ovarian hyperstimulation with expectant management (Dutch Society of Obstetrics and Gynaecology 2000; Dutch Society of Obstetrics and Gynaecology 1998; Steures *et al.,* 2006).

The follow-up started at completion of the basic fertility work-up and ended after 12 months. The primary end point was spontaneous conception resulting in an ongoing pregnancy. The first day of that menstrual cycle was considered to mark the end of time until spontaneous pregnancy. A spontaneous ongoing pregnancy was defined as the presence of fetal cardiac

activity at transvaginal sonography at a gestational age of at least 12 weeks that had resulted from treatment-independent conception.
Follow-up time ended when treatment started or at the last date at which it was sure that an ongoing pregnancy had not occurred. For all couples that were lost to follow-up, the general practitioner was sent a questionnaire on the last known reproductive status of the couple.

Analysis

We analyzed the prognostic value of a previous pregnancy by the current partnership, a previous pregnancy in a previous partnership of either partner, and previous pregnancy after fertility treatment.

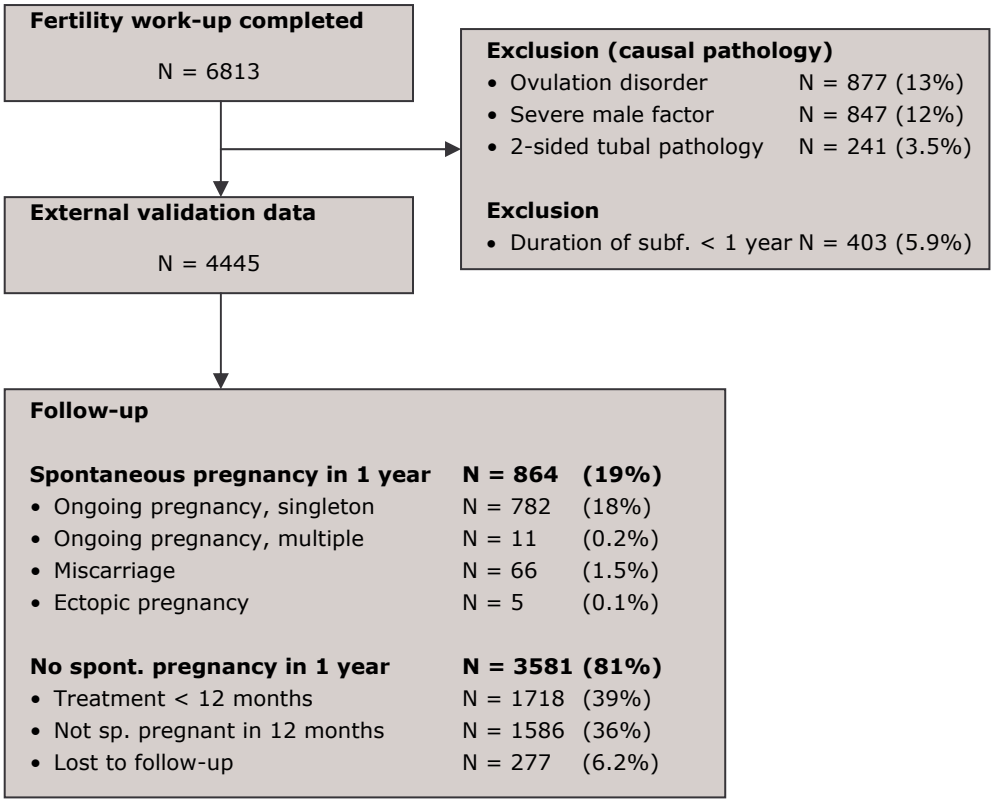
We performed both univariable and multivariable Cox' proportional hazard analyses to assess the associations between any type of previous pregnancy with time to spontaneous ongoing pregnancy as the dependent variable. For each variable, we calculated the mean and the 5^{th} to 95^{th} percentiles, but in case of a non-normal distribution we calculated the median and the 5^{th} to 95^{th} percentiles. In the multivariable analysis, we controlled for potential cofactors as female age, duration of subfertility, referral status, and semen motility. These four factors are considered as confounders because they are known predictors for spontaneous pregnancy (Hunault *et al.,* 2004). For the multivariable regression analysis we used a stepwise backwards selection procedure. Usually, the selection of variables is performed with a significance level of 5%. As the incorrect exclusion of a factor would be more deleterious than including too many factors, we selected all prognostic variables with a significance level up to 30% (Steyerberg *et al.,* 1999). Furthermore, we stratified on center level to exclude a clustering effect from the centers. Associations were expressed as hazard rate ratios (HR).

Calculations were performed with SPSS 11.0 (SPSS Inc., Chicago, IL) and S-plus 6.0 (MathSoft Inc., Seattle, WA) programs.

Results

We registered 6813 couples who completed their fertility work-up. We had to exclude 403 (6%) couples with a duration of subfertility less than one year, 877 (13%) couples with an ovulation disorder, 847 (12%) couples with severe male subfertility, and 241 (4%) couples with two-sided tubal pathology. Therefore, 4445 of the 6813 (65%) couples fulfilled the inclusion criteria. The study profile is shown in the figure 1.

Figure 1 Study profile.



Baseline characteristics are presented in table 1. The mean female age was 32.4 years, the mean male was age 35.1 years and the median duration of subfertility was 1.5 years. Subgroup analysis of the baseline characteristics for couples who had conceived before in current or a previous partnership is presented in table 2. Data on pregnancy history are summarized in table 1. Thirty percent of all couples had conceived before in their current partnership, and 9.3% of all women and 9.4% of all men had conceived in a previous partnership. Moreover, 3.2% of all couples had conceived previously after fertility treatment in their current partnership, and 0.2% of all women had conceived after fertility treatment in a previous partnership. Seventeen percent of the couples had a previous pregnancy resulting in live birth, 12% had a previous miscarriage, 1.7% had an induced abortion, and 1.2% had an ectopic pregnancy.

The follow-up status of all patients at 12 months is shown in figure 1. We completed the follow-up evaluation for 4168 couples (94%). Of all couples, 793 (18%) had a spontaneous ongoing pregnancy within one year, including 11 (1.4%) multiple pregnancies. Unsuccessful pregnancies were observed in 71 couples (1.6%), of which 66 (7.4% of all pregnancies) resulted in a miscarriage and five (0.6% of all pregnancies) resulted in an ectopic pregnancy. Within 12 months, 1718 (39%) of all couples had started treatment; 1586 (36%) did not start treatment or became pregnant.

Information on pregnancy history was missing in seven women (0.2%) with respect to their pregnancies in previous partnerships and in 183 men (4.1%) with respect to their pregnancies in previous partnerships. No data was missing on the pregnancy history of current partnerships. For one woman, the information on her age was missing, for 33 couples (0.7%) the referral status was unknown, and for 223 couples (5.0%) the data about the semen motility were missing. These missing data were imputed ("filled in") because deleting them would lead to a loss of statistical power in the multivariable analysis and -more seriously- potentially bias the results (Little and Rublin 1987; Schafer 1997). We generated a single imputed data set using the first step of the "aRegImpute" multiple imputation function in Splus 6.0. Missing data of the variables of interest and primary end point, i.e. time to spontaneous ongoing pregnancy (N = 277, 6.2%) were not imputed.

Table 1 Baseline characteristics.

	Mean	**Median**	**5th-95th percentile**
Female age (years)	32.4		25 – 39
Male age (years)	35.1		27 – 45
Duration of subfertility (years)		1.5	1.0 – 4.1
Semen analysis – TMC (10^6)		54.1	5.4 – 292
Cycle length (days)		28.1	24 – 33
Basal FSH day 2-4 (U/L)		6.5	3.4 – 12
Body Mass Index (kg/m^2)		22.9	19 – 34
Abnormal postcoital test	34%		
1-sided tubal pathology	12%		
	Total women N[a] (N = 4445)		**%**
Previous spontaneous pregnancy current partnership	1326		30
• Previous live birth	773		17
• Previous miscarriage	537		12
• Previous induced abortion	75		1.7
• Previous ectopic pregnancy	52		1.2
Previous spontaneous pregnancy woman in other partnership	412		9.3
• Previous live birth	154		3.5
• Previous miscarriage	67		1.5
• Previous induced abortion	218		4.9
• Previous ectopic pregnancy	7		0.2
Previous spontaneous pregnancy man in other partnership	417		9.4
Previous pregnancy after ART	151		3.4
• Current partnership	141		3.2
• Other partnership woman	11		0.2

For variables without a normal distribution, median is provided instead of the mean. TMC = total motile sperm count.

[a] Women who conceived more than once with different pregnancy outcomes contribute to more than one row.

Table 2 Baseline characteristics per subgroup of primary and secondary subfertility.

	N	Female age (y)	Male age (y)	Duration of subfer-tility (y)	TMC	Cycle length (d)	Basal FSH (U/L)	BMI (kg/m^2)	Abn. PCT	1-sided tubal pathology
Couple with a previous pregnancy in current partnership	1326	33.3	36.0	1.4	61.4	28.1	6.7	23.3	22%	15%
Couple with no pregnancy in current partnership, nor in previous partnership	2568	31.4	33.9	1.6	52.9	28.1	6.4	22.8	38%	9.7%
Couple with no pregnancy in current partnership, woman conceived before in previous partnership	213	34.8	35.4	1.6	49.2	28.1	7.0	22.4	33%	15%
Couple with no pregnancy in current partnership, man fathered child before in previous partnership	228	33.2	39.8	1.4	48.3	28.1	6.5	22.7	43%	17%
Couple with no pregnancy in current partnership, both man and woman had a pregnancy before in previous partnership	110	34.5	39.7	1.3	44.7	28.1	6.0	23.5	43%	22%

TMC = total motile sperm count; BMI = Body mass index, PCT = postcoital test. Means are provided for female and male age, medians for the others.

The proportional hazards analysis showed that secondary subfertile couples who had had a previous spontaneous intrauterine pregnancy had a higher probability of achieving a spontaneous ongoing pregnancy as compared with primary subfertile couples (table 3), with HR 1.3 [95% CI 1.1 – 1.6] and HR 1.4 [95% CI 1.1 – 1.6] for previous live birth and previous miscarriage in the current partnership, respectively. For previous induced abortion in the current partnership, we observed a positive probability for better fecundity (HR 1.4 [95% CI 0.91 – 2.0]). Previous spontaneous intrauterine pregnancies of the female partner in a previous partnership were not associated with fecundity. Previous spontaneous pregnancies of the male partner in a previous partnership were associated with lower fecundity (HR 0.70 [95% CI 0.54 – 0.91]). Women who had conceived after fertility treatment in the current partnership had a lower fecundity (HR 0.59 [95% CI 0.36 – 0.97]) as compared with primary subfertile women.

A previous ectopic pregnancy, be it in current or a previous partnership, was associated with a lower fecundity, although the association was not statistically significant.

These findings were confirmed in the multivariable analysis, in which we corrected for the effects of female age, duration of subfertility, referral status, and semen motility (table 3). This multivariable analysis was based on 3976 (89%) out of the 4445 couples for whom we had complete information about pregnancy history and time to spontaneous ongoing pregnancy.

Table 3 Results of the univariable and multivariable Cox' proportional hazard analysis.

	Univariable analysis		Multivariable analysis	
	HR	**95% CI**	**HR**	**95% CI**
Female age (years)			0.96	0.94–0.97
Duration of subfertility (years)			0.68	0.62–0.74
Referral status (3rd care)			0.71	0.54–0.93
Semen – progressive motility (%)			1.006	1.002–1.01
Previous spontaneous pregnancy current partnership	1.5	1.3–1.7		
• Previous live birth	• 1.3	1.1–1.6	1.4	1.2–1.7
• Previous miscarriage	• 1.4	1.1–1.6	1.3	1.0–1.5
• Previous induced abortion	• 1.4	0.91–2.0	1.4	0.90–2.1
• Previous ectopic pregnancy	• 0.73	0.40–1.3	- [a]	
Previous spontaneous pregnancy woman in other partnership	0.88	0.69–1.1		
• Previous live birth	• 0.98	0.67–1.4	- [a]	
• Previous miscarriage	• 0.69	0.37–1.3	- [a]	
• Previous induced abortion	• 1.0	0.71–1.4	- [a]	
• Previous ectopic pregnancy	• 0.63	0.01–4.5	- [a]	
Previous spontaneous pregnancy man in other partnership	0.70	0.54–0.91	0.76	0.58–0.99
Previous pregnancy after ART	0.59	0.36–0.97		
• Current partnership	• 0.64	0.39–1.1	0.52	0.30–0.90
• Other partnership woman	• 0.0	0.0-∞	- [a]	

HR = hazard ratio; CI = confidence interval. [a] variable was eliminated in the multivariable analysis (p-value > 0.30).

Discussion

This study demonstrates that a detailed pregnancy history is more important for predicting future fertility than drawing a crude distinction between primary and secondary subfertility. We found that couples who had had intrauterine pregnancies-be it an ongoing pregnancy, a miscarriage, or even an induced abortion- in their current partnership had a higher fecundity compared with primary subfertile couples. Subfertile couples who had conceived after fertility treatment had a lower fecundity compared with primary subfertile couples. Pregnancies of women in previous partnerships were not associated with fecundity, and pregnancies in previous partnerships for the male partner were negatively associated with the probability of pregnancy.

One previous study reported female secondary subfertility to be associated with a higher probability of conceiving, with a hazard ratio of 1.7 (Eimers *et al.,* 1994). However, that study did not report separately on pregnancy in the current partnership versus pregnancy in previous partnerships. Our study shows that only previous pregnancies in current partnership determine the chance to conceive. If we had pooled data on previous pregnancies of both current and previous partnerships, we would have found a statistically significant association for female secondary subfertility as well (HR 1.3 [95% CI 1.2 – 1.5]). However, this association is only due to previous pregnancies in current partnership and not to pregnancies in previous partnerships.

A remarkable finding in our study is that couples in whom the men had fathered a child in a previous partnership were less likely to conceive in comparison with primary subfertile men. We hypothesize that this is due to negative selection bias. Men who fathered a child in a previous partnership are proven to have a good fecundity; thus, one may expect these men to father a child in a new partnership without a problem except for those paired with a subfertile woman. The latter group has a higher chance of presenting at a fertility clinic, whereas the first group will conceive quickly. The fact that a couple in which the woman had conceived in a previous partnership does not have a lower fertility chance indicates that the impact of the woman on fecundity is probably much larger than that of the man.

One recent study reported that fertile women who had had a previous miscarriage had a lower fecundity compared with those who had had a previous live birth (Hassan and Killick 2005). In subfertile couples, we observed the opposite effect. Subfertile couples with a previous miscarriage

were 1.3 times more likely to conceive spontaneously. An explanation for this apparent contradiction is the fact that women with a previous miscarriage in the study of Hassan and Killick were older, more obese, and more often subfertile (OR 1.8 [95% CI 1.1 – 2.8]) (Hassan and Killick 2005). The lower fecundity after a miscarriage in that study may in fact be due to age, body weight, or other fecundity-reducing factors rather that the miscarriage itself.

Our study could not demonstrate a statistically significant lower probability of conceiving among women who had had a previous ectopic pregnancy, although a negative trend was observed for both an ectopic pregnancy in the current partnership (HR 0.73 [95% CI 0.40 – 1.3]) and an ectopic pregnancy in the previous partnership (HR 0.63 [95% CI 0.01 - 4.5]). The fact that this is not statistically significant may be due to the low prevalence of ectopic pregnancies, but, it is obvious that a previous ectopic pregnancy is an indication for tubal disease, which is likely to affect future fecundity of the couple.

Our study demonstrates that couples with a previous spontaneous intrauterine pregnancy in their current partnership have a better fecundity than couples with primary subfertility. Couples with a previous pregnancy after fertility treatment and those with a pregnancy by the male partner in a previous partnership have lower spontaneous pregnancy chances. To predict the occurrence of spontaneous ongoing pregnancy in subfertile couples, it is important to know the exact fertility history of the couple, including the outcome, localization, and origin and mode of conception of a previous pregnancy rather than just categorizing couples as primary or secondary subfertile.

CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Investigators, participating centers in the Netherlands:
Y.M. van Kasteren; Alkmaar, Medisch Centrum Alkmaar; P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis; W.A. Schöls; Amersfoort, Meander Medisch Centrum; M.H. Mochtar; Amsterdam, Academisch Medisch Centrum; G.L.M. Lips; Amsterdam, BovenIJ Ziekenhuis; H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis; S. Milosavljevic; Amsterdam, Slotervaart Ziekenhuis; P.G.A. Hompes; Amsterdam, Vrij Universiteit Medisch Centrum; A.V. Sluijmer; Assen, Wilhelmina Ziekenhuis; H.E. Bobeck; Beverwijk, Rode Kruis Ziekenhuis ; R.E. Bernardus; Blaricum, Tergooiziekenhuizen; J.P. Dörr; Den Haag, Westeinde Ziekenhuis; H.J.M. Roelofs; Dordrecht, A. Schweitzer Ziekenhuis; J.M. Burggraaff; Emmen, Scheper Ziekenhuis; G.J.E. Oosterhuis; Enschede, Medisch Spectrum Twente; M.H. Schouwink; Geldrop, St. Anna Ziekenhuis; P.X.J.M. Bouckaert; Heerlen, Atrium Medisch Centrum; F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis; M. van Hoven; Hilversum, Tergooiziekenhuizen; M.H. Emanuel; Hoofddorp, Spaarne Ziekenhuis; C.N.M. Renckens; Hoorn, Westfries Gasthuis; J.A. Land; Maastricht, Academisch Ziekenhuis Maastricht; J.H. Schagen-Van Leeuwen; Nieuwegein, St. Antonius Ziekenhuis; J.A.M. Kremer; Nijmegen, UMC St. Radboud; C. van Katwijk; Purmerend, Waterland Ziekenhuis; H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis; H.J.L.A. Ruis; Veghel, Ziekenhuis Bernhoven; C.A.M. Koks; Veldhoven, Máxima Medisch Centrum; P. Bourdrez; Venlo/Venray, Vie Curi Medisch Centrum; W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum

References

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Dutch Society of Obstetrics and Gynaecology. (1998) Guideline-Indications for invitro fertilization (IVF). *NVOG-richlijn nr 9,*1-7.

Dutch Society of Obstetrics and Gynaecology. (2000) Guideline-Intra-uterine insemination (IUI). *NVOG-richtlijn nr 29,*1-8.

Dutch Society of Obstetrics and Gynaecology. (2004) Guidline-Basic fertility work-up. *NVOG-richtlijn nr 1,*1-12.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, and Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18,1959-1966.

Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, and Freundl G. (2005) Definition and prevalence of subfertility and infertility. *Hum Reprod* 20,1144-1147.

Hassan MA and Killick SR. (2005) Is previous aberrant reproductive outcome predictive of subsequently reduced fecundity? *Hum Reprod* 20,657-664.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, and Desai KM. (1985) Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 291,1693-1697.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Little,R.J.A. and Rublin,D.B. (1987) Statistical analysis with missing data. Wiley, New York.

Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Van der Veen F, and Bossuyt PMM. (1997) The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 67,1031-1037.

Schafer,J.L. (1997) Analysis of incomplete multivariate data. Chapman & Hall, London.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Steures P, Van der Steeg JW, Hompes PGA, Habbema JDF, Eijkemans MJC, Broekmans FJ, Verhoeve HR, Bossuyt PM, Van der Veen F, and Mol BWJ. (2006) Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 368,216-221.

Steyerberg EW, Eijkemans MJC, and Habbema JDF. (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 52,935-942.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Van Dessel HJ, Bossuyt PM, Van der Veen F, and Mol BW. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,536-542.

Wang X, Chen C, Wang L, Chen D, Guang W, and French J. (2003) Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79,577-584.

HUMAN REPRODUCTION
2008; 23: 324 - 328

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Peter G.A. Hompes
Jan M. Burggraaff
G. Jur E. Oosterhuis
Patrick M.M. Bossuyt
Fulco van der Veen
Ben Willem J. Mol
for the CECERM investigators

chapter (07)

OBESITY AFFECTS SPONTANEOUS PREGNANCY

chances in subfertile, ovulatory women

Abstract

Background Obesity is increasing rapidly among women all over the world. Obesity is a known risk factor for subfertility due to anovulation, but it is unknown whether obesity also affects spontaneous pregnancy chances in ovulatory subfertile women.
Methods We evaluated whether obesity affected the chance of a spontaneous pregnancy in a prospectively assembled cohort of 3029 consecutive subfertile couples. Women had to be ovulatory and had to have at least one patent tube, whereas men had to have a normal semen analysis. Time to spontaneous ongoing pregnancy within twelve months was the primary endpoint.
Results The probability of a spontaneous pregnancy declined linearly with a body mass index (BMI) over 29 kg/m^2. Corrected for possible related factors, women with a high BMI had a 4% lower pregnancy rate per kg/m^2 increase (hazard ratio: 0.96 [95% CI 0.91-0.99]).
Conclusions These results indicate that obesity was associated with lower pregnancy rates in subfertile ovulatory women.

Introduction

The use of assisted fertility techniques has increased tremendously in the past three decades (Society for Assisted Reproductive Technology and American Society for Reproductive Medicine 2002). Although this may be due to better availability of the wide range of current technologies, an increased demand for fertility care may play a role as well. This increased demand may be due to an increased incidence of Chlamydia trachomatis and increased maternal age (Coombes 2004; Martin 2000). Additionally, obesity is expected to be a potential cause for an increase in subfertility in the near future (Bolúmar *et al.,* 2000).

Obesity is increasing rapidly all over the world, affecting more than one billion people worldwide (Haslam and James 2005). The World Health Organisation (WHO) considers a body mass index (BMI) as abnormal if BMI is over 25.0 kg/m^2 and defines obesity as a BMI over 30.0 kg/m^2 (World Health Organisation 1995). More women of reproductive age become overweight and obese. Nowadays, the incidence of obesity in women of child bearing age is 12% in Western Europe and 25% in North America (Butler 2004; Linné 2004; Watson 2005; Haslam and James 2005). The main adverse consequences are cardiovascular disease, type 2 diabetes, and cancer. Overall it is thought to be the sixth most important risk factor for mortality and morbidity (Allison *et al.,* 1999). Furthermore, obesity is a known risk factor for anovulation, which may lead to subfertility (Rogers and Mitchell, Jr. 1952; Norman and Clark 1998; Hartz *et al.,* 1979; Moran *et al.,* 1999).

Current NICE fertility guidelines recommend that all obese women, regardless of their cycle characteristics, should be informed that they are likely to take longer to conceive (National institute for clinical excellence 2004). This recommendation is based on three studies. Two studies analysed the relationship between body mass index and time to pregnancy in women who were pregnant or had delivered a child (Bolúmar *et al.,* 2000; Jensen *et al.,* 1999), whereas the third study analysed fat distribution and the chance of conceiving in women in a donor insemination programme (Zaadstra *et al.,* 1993). All studies reported a negative effect of obesity on the chance of pregnancy in these potentially fertile women.

However, evidence that obesity also affects the chance of spontaneous pregnancy, in subfertile ovulatory women is still lacking. The aim of this study was to determine whether obesity in subfertile ovulatory women is associated with a decreased chance of spontaneous pregnancy.

Materials and methods

Between January 2002 and February 2004, we included consecutive subfertile couples that had not been evaluated previously for subfertility, in a prospective cohort study. The study was performed in 24 hospitals in The Netherlands. The detailed study protocol has been documented in a previous publication (Van der Steeg *et al.,* 2007).

In short, all couples underwent a fertility work-up consisting of: a fertility history, including details about height and weight, smoking habits, assessment of ovulation, assessment of tubal patency and semen analysis (Dutch Society of Obstetrics and Gynaecology 2004). Duration of subfertility and female and male age were set at the end of the infertility assessment. Subfertility was considered to be secondary if a woman had conceived in the current or in a prior partnership, regardless of the pregnancy outcome. The BMI was calculated as the weight in kilograms divided by the square of the height in meters, both self-reported during the first visit. BMI of the men as well as timing and frequency of intercourse were not documented.

Fertility work-up of the female partner

Ovulation was assessed by means of a basal body temperature chart, measurement of mid-luteal serum progesterone, or by sonographic monitoring of the cycle. The menstrual cycle was considered regular if the duration of the cycle was between 23 and 35 days, with an intercycle variation of less than eight days, during the past year (Munster *et al.,* 1992). Tubal pathology was assessed by a chlamydia antibody test (CAT), hysterosalpingography (HSG) or laparoscopy. Those with a positive CAT subsequently went on to have further investigation with a HSG or laparoscopy (Mol *et al.,* 1997). Couples, in whom the female partner was diagnosed with anovulation or with two-sided tubal pathology, were excluded from the analysis.

Fertility work-up of the male partner

Semen analysis was performed at least once according to the WHO guidelines, including semen volume, concentration, morphology and motility (World Health Organization 1999). Couples in whom the man had a total motile sperm count (TMC) $< 3 \times 10^6$ were excluded from the study.

Follow-up

After completion of the fertility work-up the probability of a spontaneous pregnancy within one year, leading to live birth, was calculated with a validated prediction model (www.freya.nl/probability.php) (Hunault *et al.,* 2004; Van der Steeg *et al.,* 2007). Depending on that probability couples were counselled for expectant management or fertility treatment according to the national fertility guidelines (Dutch Society of Obstetrics and Gynaecology 1998; Dutch Society of Obstetrics and Gynaecology 2000). The exact study flow has been reported in a previous paper (Van der Steeg *et al.,* 2007). Couples were followed prospectively from the completion of the fertility work-up until pregnancy or start of treatment within 12 months. The primary endpoint was time to conception without treatment, resulting in an ongoing pregnancy, counted in calendar time used in a continuous way. Couples who did not conceive were censored when treatment started or at the last date of contact during follow-up.

Analysis

We first assessed the relation between BMI and probability of pregnancy through spline functions. By visual inspection, it was determined whether BMI behaved as a linear or non-linear function in relation to the probability of spontaneous pregnancy, and whether cut-off values for optimal BMIs could be observed. Non-linearity was tested with ANOVA analysis (Harrell, Jr. *et al.,* 1988).

We then analysed the predictive capacity of BMI as hazard ratios (HR) by Cox proportional hazard analysis of the time to spontaneous ongoing pregnancy. The proportional hazards assumption was evaluation with S-plus (Grambsch and Therneau 1994). Lastly, we repeated the analysis correcting for possibly related factors in a multivariable hazards regression model. Potential related factors were female age, duration of subfertility, previous pregnancy, referral status, semen motility and current smoking of the female and male partner (Hunault *et al.,* 2004; Bolúmar *et al.,* 2000).

In all analyses, a *p*-value of 0.05 was used to indicate significance. Calculations were performed with SPSS® 12.0 (SPSS Inc., Chicago, IL) and S-plus® 6.0 (MathSoft Inc., Seattle, WA) programs.

Results

In total, 6035 subfertile couples that had completed their fertility work-up were registered. Of these, 379 (6.3%) couples with a duration of subfertility less than one year, 692 (12%) couples with anovulation, 211 (3.5%) couples with two-sided tubal pathology and 699 (12%) couples with severe male factor were excluded (figure 1). In 1025 (17%) couples the BMI was not reported. Therefore, 3029 couples were included in the analysis. Follow-up was completed for 2793 couples (92%). Of all couples, 529 (17%) had a spontaneous ongoing pregnancy within one year (figure 1). In 17 women, pregnancy outcome was unknown. There were 47 women (7.8% of all pregnancies) who miscarried and four women (0.7% of all pregnancies) who had an ectopic pregnancy. Within 12 months, 1136 (38%) started treatment, whereas 1060 (35%) neither started treatment nor became pregnant. Median length of follow-up was 28.1 weeks (5^{th} to 95^{th} percentile: 2 to 134 weeks), 18.5 weeks (1 to 100 weeks) for those who conceived and 31.5 weeks (2 to 146 weeks) for those without a pregnancy.

Baseline characteristics are represented in the table 1. The median BMI was 22.9 kg/m^2 (5^{th} to 95^{th} percentile: 19-33 kg/m^2). A BMI below 18.5 kg/m^2 was found in 3.7% of the women, between 18.5 and 25 in 67%, between 25 and 30 in 19%, between 30 and 35 kg/m^2 in 6.7%, and ≥ 35 kg/m^2 in 3.8%. Couples, in whom the BMI was not documented, were on average older, more often secondary subfertile and more often referred by a gynaecologist (ANOVA statistics, $P < 0.05$), although differences were small. Other baseline characteristics were comparable between the groups.

Figure 1 Flow chart.

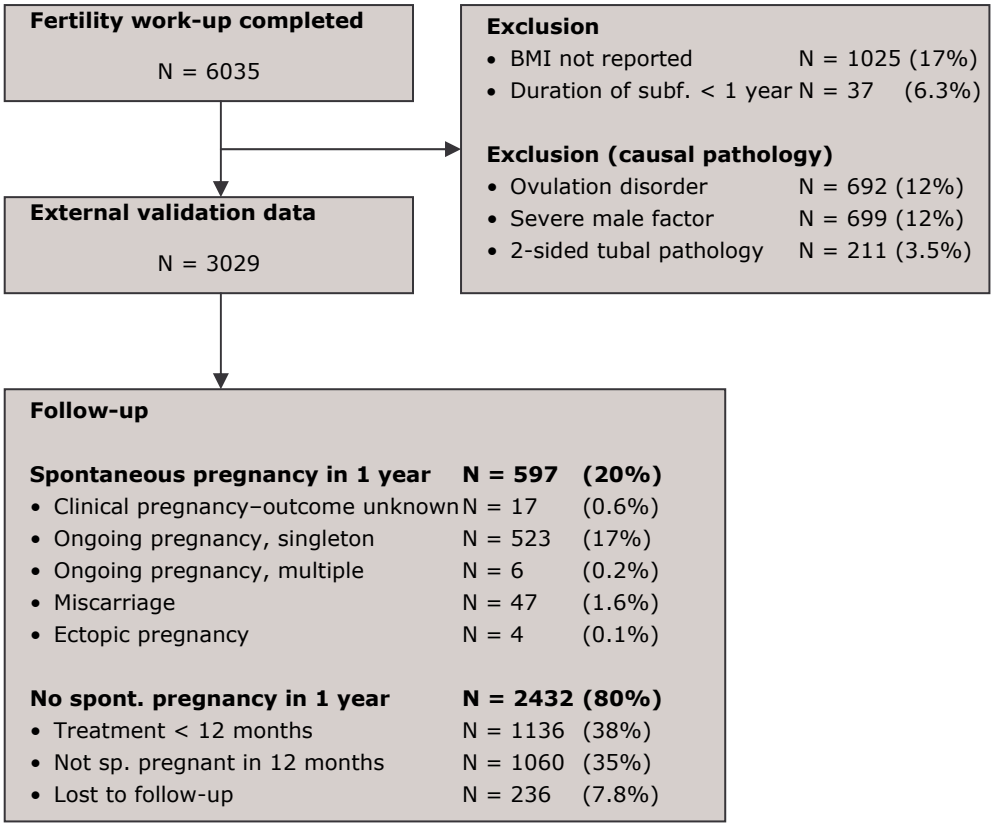


Table 1 Baseline characteristics.

	BMI available N = 3029		BMI missing N = 1025		
	Mean	**5th-95th percentile**	**Mean**	**5th-95th percentile**	***p*-value[a]**
Female age (y)	32.1	25 – 39	32.9	26 – 40	< 0.01
Male age (y)	34.9	27 – 44	35.7	28 – 46	< 0.01
Duration of subfertility (y) (median)	1.5	1.0 – 4.0	1.5	1.0 – 4.3	0.41
BMI (kg/m^2)	22.9	19 – 33	-	-	-
Subfertility, primary (%)	73		66		< 0.01
Subfertility, secondary (%)	27		34		-
Referral status (2^{nd} care)	92		86		< 0.01
Referral status (3^{rd} care)	7.8		14.1		-
Semen– TMC (10^6) (median)	51.0	5.3 – 284	50.8	6.0 – 290	0.22
Cycle length (days)	28.1	23 – 33	28.1	23 – 33	0.84
Current smoking woman, no (N,%)	2272	(75%)	2820	(80%)	0.03
Current smoking woman, yes (N,%)	757	(25%)	209	(20%)	-
Current smoking man, no (N,%)	20 05	(66%)	328	(68%)	0.29
Current smoking man, yes (N,%)	1024	(34%)	2701	(32%)	-

[a] Difference in baseline characteristics between women with and without data on BMI, tested with ANOVA.

The spline analysis showed that BMI had an inversed U-shaped relationship with the probability of pregnancy, although this was not statistically significant over the whole range (figure 2) (ANOVA: p = 0.4). From this spline function two thresholds were derived at 21 and 29 kg/m^2. Women with a BMI between 21 and 29 were defined as the reference group.

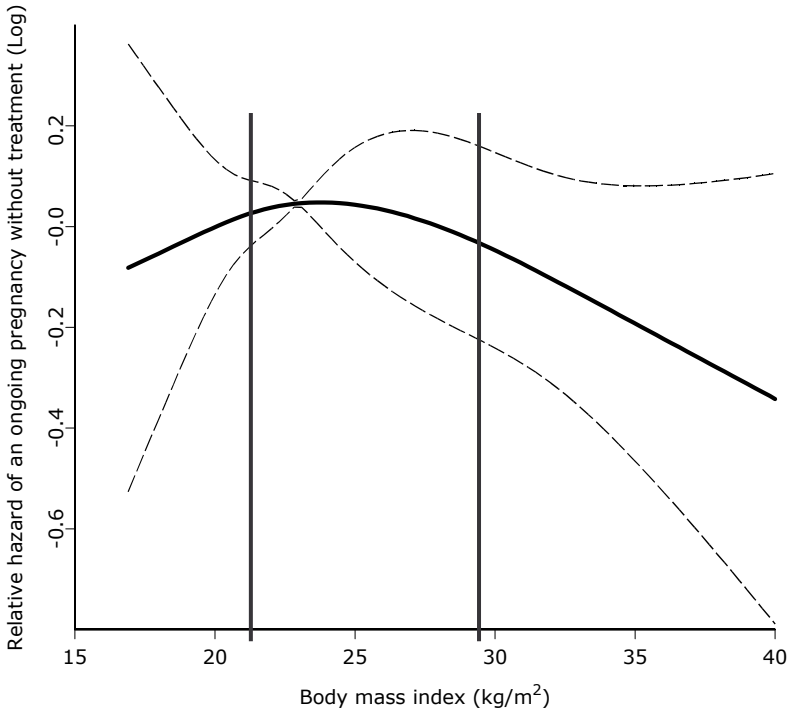


Figure 2 Spline function of the BMI in relation to time to spontaneous ongoing pregnancy.
Relative hazard ~ HR. Dotted lines represent 95% confidence intervals. Vertical lines show the thresholds of 21 and 29 kg/m^2. BMIs above 29 kg/m^2 were significantly associated with a decreased fecundity, whereas there was a trend below 21 kg/m^2.

The Univariable analysis showed that BMI, female age, duration of subfertility, secondary subfertility, referral status and semen motility were

statistically significantly related to the probability of a spontaneous ongoing pregnancy (table 2). A BMI above 29 kg/m^2 was associated with a statistically significant lower probability of spontaneous ongoing pregnancy than the reference group (HR 0.95 per kg/m^2 above 29 kg/m^2, [95% CI 0.91 – 0.99]). A BMI below 21 kg/m^2 was associated with a lower probability of spontaneous ongoing pregnancy than the reference group, but was not statistically significant (HR 0.97 per kg/m^2 below 21 kg/m^2, [95% CI 0.87 – 1.07]). (table 2). The proportional hazards assumption, necessary to perform the Cox analysis in a correct way, was fulfilled.

Table 2 Results of the univariable and multivariable Cox' regression analysis.

	Univariable analysis		**Multivariable analysis**	
	HR	**95% CI**	**HR**	**95% CI**
BMI (kg/m^2) per unit < 21	0.97	(0.87-1.07)	0.95	(0.86-1.05)
21 – 29[a]	1	-	1	-
per unit ≥ 29	0.95	(0.91-0.99)	0.96	(0.91-0.99)
Potential confounders				
Female age (y) < 31 per y	0.99	(0.95-1.02)	0.98	(0.94-1.01)
≥ 31 per y	0.94	(0.91-0.98)	0.94	(0.90-0.97)
Duration of subfertility per y	0.91	(0.84-0.99)	0.94	(0.86-1.03)
Subfertility, primary	1	-	1	-
Subfertility, secondary	1.5	(1.3-1.8)	1.6	(1.3 -1.9)
Referral status, 2nd care	1	-	1	-
Referral status, 3rd care	0.65	(0.46-0.92)	0.67	(0.48-0.95)
Semen motility (per %)	1.009	(1.004-1.013)	1.008	(1.004-1.013)
Current smoking woman, no	1	-	1	-
Current smoking woman, yes	1.1	(0.96-1.4)	1.2	(0.98-1.4)
Current smoking man, no	1	-	1	-
Current smoking man, yes	0.97	(0.82-1.1)	0.94	(0.78-1.1)

HR = hazard ratio; CI = confidence interval. [a] Women with a BMI 21 – 29 kg/m^2 were used as reference group.

The multivariable analysis adjusted for female age, duration of subfertility, previous pregnancy, referral status (secondary or tertiary care), semen motility, and current smoking of the female and male partner, and did not change the results (HR 0.96 per kg/m^2, [95% CI 0.91 – 0.99]). In case of a woman with a BMI of 35 kg/m^2, the probability of spontaneous pregnancy was 26% lower, and in case of a woman with a BMI of 40 kg/m^2 it was 43% lower compared with women with a BMI between 21 and 29 kg/m^2.

Discussion

This cohort study showed that obesity is an important risk factor for pregnancy chances in ovulatory, subfertile women. For every BMI unit above 29 kg/m^2 the probability was reduced by approximately 5%, being a reduction comparable with the increment of one year in female age. Given the increased prevalence of obesity this is a worrying finding.

Up till now, the relationship between BMI and pregnancy chances had not been established in ovulatory subfertile women. This is the first prospective cohort study to demonstrate this. It differs from previous studies on obesity and pregnancy chances in two ways. First, all other studies dealt with proven fertile populations, whereas our study included subfertile women. Second, many studies dealt with obesity as a categorical variable. In contrast, in this study, BMI was analysed as a continuous variable that allowed subtle decline in pregnancy rate starting at 29 kg/m^2 to be demonstrated. In proven fertile women, BMI was reported to be a risk factor for the chance of conception in the category of women with a BMI over 25 kg/m^2 (HR 0.77 (95% CI 0.70 – 0.84) (Jensen *et al.,* 1999). In the category of women with a BMI over 30 kg/m^2, BMI was reported to be a risk factor of having a delayed conception with an odds ratio of 12 (95% CI 3.7 – 36) (Bolúmar *et al.,* 2000).

A limitation of this study is that frequency of intercourse was not taken into account. Recently, a review found support that obesity is associated with decreased intercourse frequency, reduced sexual desire and erectile dysfunction (Larsen *et al.,* 2007). However, in view of the paucity of data, confounding factors like medication and adverse lifestyles could not be ruled out.

Another limitation of our study is the fact that the BMI of the male partner was not taken into account. Male obesity has been reported to increase the chance of becoming subfertile (Ramlau-Hansen *et al.,* 2007), although the effect was weak. Nevertheless, because these data were missing in our study, we were not able to confirm or reject their findings. Finally, in the present study, BMI was lacking in 17% of the women. As our purpose was to examine the relation between BMI and the probability of pregnancy, this could have led to biased estimates of associations. To examine the impact of this partial verification, we explored whether there were any systematic differences between women in whom BMI was known and women in whom data on BMI was lacking. This was indeed the case with respect to female age, duration of subfertility, being secondary

subfertile, referral status and smoking habits of the women, although differences were small. However, we may have selected a group here that was more fertile than the overall group.

We can only speculate about the pathophysiological explanations for the lower pregnancy chances in obese women. It has been suggested that leptin may be of importance (Mantzoros 2000; Chan and Mantzoros 2005; Rosenbaum and Leibel 1999). Genetically mediated states of leptin deficiency result in obesity and subfertility (Rosenbaum and Leibel 1999). Decreasing leptin levels due to starvation result in decreased estradiol levels and amenorrhoea (Mantzoros 2000). There is evidence that leptin may influence ovarian steroidogenesis directly. Further research of the role of intra-ovarian leptin action in relation to subfertility remains of interest.

It could be hypothesized that lifestyle interventions that focus on weight reduction are an effective intervention (Knowler *et al.,* 2002). This study focused on the BMI at start of the fertility work-up, rather than on weight changes. In subfertile anovulatory women several studies reported such a beneficial effect (Hollmann *et al.,* 1996; Pasquali *et al.,* 1997; Crosignani *et al.,* 2003). A next step could be to randomly allocate obese and ovulatory subfertile women to a controlled low-calories diet, or to their normal diet and compare hazard ratios.

In conclusion, the results of this study indicate that ovulatory subfertile women with a BMI over 29 kg/m^2 have lower pregnancy rates compared with those with normal weight. Now, we know that not only obese women with anovulation have lower chances of conception, but also obese women with a regular cycle. Due to the fact that more women of child-bearing age become overweight and obese, this is a worrying finding.

Acknowledgements

The authors thank all participating hospitals for their contribution to this study. Furthermore, Dr A. Bell, MB. ChB. MRCPCH. Clinical Research Fellow Paediatric Cardiology (Division of Imaging Sciences, Kings College, London) is acknowledged for his critical reading and valuable suggestions on previous drafts of this paper.

CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Investigators, participating centres in The Netherlands:
Y.M. van Kasteren; Alkmaar, Medisch Centrum Alkmaar; P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis; W.A. Schöls; Amersfoort, Meander Medisch Centrum; M.H. Mochtar; Amsterdam, Academisch Medisch Centrum; G.L.M. Lips; Amsterdam, BovenIJ Ziekenhuis; H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis; S. Milosavljevic; Amsterdam, Slotervaart Ziekenhuis; P.G.A. Hompes; Amsterdam, Vrij Universiteit Medisch Centrum; J.P. Dörr; Den Haag, Westeinde Ziekenhuis; P.J.Q. van der Linden; Deventer, Ziekenhuis Deventer; H.J.M. Roelofs; Dordrecht, A. Schweitzer Ziekenhuis; J.M. Burggraaff; Emmen, Scheper Ziekenhuis; G.J.E. Oosterhuis; Enschede, Medisch Spectrum Twente; M.H. Schouwink; Geldrop, St. Anna Ziekenhuis; P.X.J.M. Bouckaert; Heerlen, Atrium Medisch Centrum; F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis; C.J.C.M. Hamilton; 's Hertogenbosch, Jeroen Bosch Ziekenhuis; J.A. Land; Maastricht, Academisch Ziekenhuis Maastricht; J.H. Schagen-Van Leeuwen; Nieuwegein, St. Antonius Ziekenhuis; J.A.M. Kremer; Nijmegen, UMC St. Radboud; H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis; C.A.M. Koks; Veldhoven, Máxima Medisch Centrum; P. Bourdrez; Venlo/Venray, Vie Curi Medisch Centrum; W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum

References

Allison DB, Fontaine KR, Manson JE, Stevens J, and VanItallie TB. (1999) Annual deaths attributable to obesity in the United States. *JAMA* 282,1530-1538.

Bolúmar F, Olsen J, Rebagliato M, Saez-Lloret I, and Bisanti L. (2000) Body mass index and delayed conception: a European Multicenter Study on Infertility and Subfecundity. *Am J Epidemiol* 151,1072-1079.

Butler D. (2004) The fertility riddle. *Nature* 432,38-39.

Chan JL and Mantzoros CS. (2005) Role of leptin in energy-deprivation states: normal human physiology and clinical implications for hypothalamic amenorrhoea and anorexia nervosa. *Lancet* 366,74-85.

Coombes R. (2004) Doctors demand national screening for chlamydia. *BMJ* 328,1397.

Crosignani PG, Colombo M, Vegetti W, Somigliana E, Gessati A, and Ragni G. (2003) Overweight and obese anovulatory patients with polycystic ovaries: parallel improvements in anthropometric indices, ovarian physiology and fertility rate induced by diet. *Hum Reprod* 18,1928-1932.

Dutch Society of Obstetrics and Gynaecology. (1998) Guideline-Indications for invitro fertilization (IVF). *NVOG-richlijn nr 9,*1-7.

Dutch Society of Obstetrics and Gynaecology. (2000) Guideline-Intra-uterine insemination (IUI). *NVOG-richtlijn nr 29,*1-8.

Dutch Society of Obstetrics and Gynaecology. (2004) Guideline-Basic fertility work-up. *NVOG-richtlijn nr 1,*1-12.

Grambsch PM and Therneau TM. (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81,515-526.

Harrell FE, Jr., Lee KL, and Pollock BG. (1988) Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 80,1198-1202.

Hartz AJ, Barboriak PN, Wong A, Katayama KP, and Rimm AA. (1979) The association of obesity with infertility and related menstural abnormalities in women. *Int J Obes* 3,57-73.

Haslam DW and James WP. (2005) Obesity. *Lancet* 366,1197-1209.

Hollmann M, Runnebaum B, and Gerhard I. (1996) Effects of weight loss on the hormonal profile in obese, infertile women. *Hum Reprod* 11,1884-1891.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Jensen TK, Scheike T, Keiding N, Schaumburg I, and Grandjean P. (1999) Fecundability in relation to body mass and menstrual cycle patterns. *Epidemiology* 10,422-428.

Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, and Nathan DM. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346,393-403.

Linné Y. (2004) Effects of obesity on women's reproduction and complications during pregnancy. *Obes Rev* 5,137-143.

Mantzoros CS. (2000) Role of leptin in reproduction. *Ann N Y Acad Sci* 900,174-183.

Martin SP. (2000) Diverging fertility among U.S. women who delay childbearing past age 30. *Demography* 37,523-533.

Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Van der Veen F, and Bossuyt PMM. (1997) The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 67,1031-1037.

Moran C, Hernandez E, Ruiz JE, Fonseca ME, Bermudez JA, and Zarate A. (1999) Upper body obesity and hyperinsulinemia are associated with anovulation. *Gynecol Obstet Invest* 47,1-5.

Munster K, Schmidt L, and Helm P. (1992) Length and variation in the menstrual cycle--a cross-sectional study from a Danish county. *Br J Obstet Gynaecol* 99,422-429.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems.
www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Norman RJ and Clark AM. (1998) Obesity and reproductive disorders: a review. *Reprod Fertil Dev* 10,55-63.

Pasquali R, Casimirri F, and Vicennati V. (1997) Weight control and its beneficial effect on fertility in women with obesity and polycystic ovary syndrome. *Hum Reprod* 12 Suppl 1,82-87.

Ramlau-Hansen CH, Thulstrup AM, Nohr EA, Bonde JP, Sorensen TI, and Olsen J. (2007) Subfecundity in overweight and obese couples. *Hum Reprod* 22,1634-1637.

Rogers J and Mitchell GW, Jr. (1952) The relation of obesity to menstrual disturbances. *N Engl J Med* 247,53-55.

Rosenbaum M and Leibel RL. (1999) The role of leptin in human physiology. *N Engl J Med* 341,913-915.

Society for Assisted Reproductive Technology and American Society for Reproductive Medicine. (2002) Assisted reproductive technology in the United States: 1998 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. *Fertil Steril* 77,18-31.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Van Dessel HJ, Bossuyt PM, Van der Veen F, and Mol BW. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,536-542.

Watson R. (2005) EU aims to tackle growing problem of obesity. *BMJ* 331,1426.

World Health Organisation. (1995) Physical Status: The Use and Interpretation of Anthropometry. *World Health Organ Tech Rep Ser* 854.

World Health Organization (1999) WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. Cambridge University Press, Cambridge, England.

Zaadstra BM, Seidell JC, Van Noord PA, Te Velde ER, Habbema JD, Vrieswijk B, and Karbaat J. (1993) Fat and female fecundity: prospective study of effect of body fat distribution on conception rates. *BMJ* 306,484-487.

THE JOURNAL OF CLINICAL ENDOCRINOLOGY & METABOLISM
2007; 92: 2163 - 68

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Peter G.A. Hompes
Frank J. Broekmans
Peter X.J.M. Bouckaert
Patrick M.M. Bossuyt
Fulco van der Veen
Ben Willem J. Mol
for the CECERM investigators

chapter (08)

PREDICTIVE VALUE AND CLINICAL IMPACT

of basal Follicle-Stimulating Hormone in subfertile, ovulatory women

Abstract

Context Basal FSH is a marker for ovarian reserve.
Objectives The objective of the study was to investigate the predictive value of basal FSH on spontaneous ongoing pregnancy in subfertile ovulatory women.
Design This was a prospective cohort study.
Setting The study was conducted in 19 fertility centers in The Netherlands.
Participants Subfertile ovulatory women without two-sided tubal pathology and in whom the man had normal sperm parameters (total motile count ≥ 3×10^6) participated in the study.
Interventions Interventions included a fertility work-up, including a basal FSH measurement on cycle day 3.
Main Outcome Measures Spontaneous ongoing pregnancy was measured.
Results We included 3519 consecutive couples of which 562 (16%) had a spontaneous ongoing pregnancy within 1 year. Basal FSH levels of 8 IU/L or higher were associated with a decreased probability of spontaneous ongoing pregnancy (hazard ratio (HR) 0.93 per IU/L [95% confidence interval (CI) 0.87 – 0.98]). In a multivariable analysis, female age (HR 0.97/y, 95% CI 0.95 – 0.99), cycle length (HR 0.96 per day, 95% CI 0.93 – 1.0), and FSH levels 8 IU/L or greater (HR 0.93 per IU/L, 95% CI 0.87 – 0.99) were strong negative predictors for spontaneous ongoing pregnancy.
Addition of FSH to a prediction model based on female age, duration of subfertility, previous pregnancy, referral status and semen analysis changed the probability to conceive spontaneously from 30% or greater to less than 30% in 97 of 3219 couples (3.0%).
Conclusions In ovulatory women, a basal FSH level of 8 IU/L or higher is associated with decreasing fecundity, independent of female age and cycle length. Because the number of couples in whom the FSH level alters management decisions is low, we do not recommend routine testing of basal FSH in subfertile couples.

Introduction

Increasing female age is generally accepted as an important factor involved in decreased fecundity (Evers 2002). It indirectly reflects the decline of ovarian reserve caused by aging (Noord-Zaadstra *et al.,* 1991; Collins and Rowe 1989). In the developed world, there is ample evidence that women postpone childbearing, a trend that can be found in some parts of Asia and Latin America as well (Bongaarts 1999; Ford and Nault 1996; Heck *et al.,* 1997). Currently one of three women has not conceived by the age of 30 years (Martin 2000). As a result of this delay in childbearing, the number of women presenting at fertility centers with age-related subfertility may increase.

In the fertility work-up, several tests have been proposed to assess a woman's individual ovarian reserve, of which measurement of basal FSH in the early follicular phase has been the most popular one (Scott, Jr. and Hofmann 1995; Sharara *et al.,* 1998). A meta-analysis of the predictive performance of basal FSH as a test for ovarian reserve in *in vitro* fertilization (IVF) patients showed that basal FSH had only a limited capacity in predicting IVF outcome. It was recommended that basal FSH should not be performed as a routine test for the prediction of IVF outcome (Bancsi *et al.,* 2003; Hendriks *et al.,* 2005c; Broekmans *et al.,* 2006).

In contrast to IVF, the predictive value of basal FSH screening for spontaneous pregnancy in a general subfertile population has not been assessed extensively. The existing knowledge on the capacity of FSH to predict spontaneous pregnancy is based on two retrospective studies and one small prospective study. These studies failed to demonstrate a clinically relevant predictive value of FSH for spontaneous ongoing pregnancy (Van Montfrans *et al.,* 2000; Van Rooij *et al.,* 2004; Van Rooij *et al.,* 2006; Scott *et al.,* 1993). Nevertheless, many clinicians use basal FSH screening in the initial evaluation of patients with subfertility (Broekmans *et al.,* 1998).

In view of this issue, we performed a prospective cohort study in a general subfertile population to assess the predictive value of basal FSH for spontaneous ongoing pregnancy in ovulatory women.

Subjects and methods

Between January 2002 and February 2004, we included consecutive subfertile couples in a prospective cohort study, performed in 19 hospitals in The Netherlands. The local ethics committee of each participating center gave institutional review board approval for this study. All couples underwent a basic fertility work-up, consisting of a fertility history, semen analysis, assessment of ovulation, basal FSH level testing on cycle day 3, postcoital test and assessment of the fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and Gynecology (Forti and Krausz 1998; Dutch Society of Obstetrics and Gynaecology 2004).

The menstrual cycle was considered regular if the duration of the cycle was between 23 and 35 d, with an intercycle variation of less than 8 d. Ovulation was assessed by means of a basal body temperature chart, a midluteal serum progesterone, or sonographic monitoring of the cycle. Women with an irregular menstrual cycle or documented anovulation were excluded.

Basal FSH concentrations were measured at least once on cycle day 2-4 in each participating center with commercially available immunometric assays (table 1).

Table 1 FSH assays with their analytical variations used in the 19 hospitals.

FSH assay	Firm	Analytical variation	Hospi-tals N	Women N (%)
Access 2	Beckman Coulter Access (Fullerton, CA)	5.6% at 9.95 IU/L 5.4% at 15.45 IU/L 4.3% at 36.40 IU/L	2	298 (8.5)
Advia Centaur	Bayer Diagnostics B.V. (Tarrytown, NY)	3.9% at 6.9 IU/L 2.4% at 23.4 IU/L 2.2% at 53.1 IU/L	4	1191 (34)
Architect platform	Abbott Laboratories (Abbott Park, JL)	4.5% at 4.2 IU/L 3.8% at 16 IU/L	1	228 (6.5)
AutoDelfia	Wallac Oy (Turku, Finland)	3.1% at 2.56 IU/L 2.1% at 11.4 IU/L 2.6% at 44.5 IU/L	2	367 (10)
AxSym FSH	Abbott Laboratories	5.8% at 3.0 IU/L 6.9% at 54 IU/L	1	217 (6.2)
Elecsys 2010	Roche (Copenhagen, Denmark)	2.6% at 6.1 IU/L 3.5% at 45.0 IU/L	3	284 (8.1)
Immulite 2000	Diagnostic Products Corp. (Los Angeles, CA)	3.1% at 5.8 IU/L 3.2% at 16.8 IU/L 2.7% at 40.9 IU/L	5	833 (24)
Modular E170	Roche	4.4% at 3.5 IU/L 4.4% at 12.5 IU/L	1	101 (2.9)
		Total	19	3519

Semen analysis was performed at least once according to the World Health Organization manual (World Health Organization 1999). The total motile sperm count (TMC) was calculated from semen volume, sperm concentration, and percentage progressive motility. Men with a prewash TMC less than 3×10^6 were not included in the study (Van Weert *et al.,* 2005).

Tubal pathology was assessed by a chlamydia antibody test (CAT), or directly assessed by HSG or laparoscopy.

In case of a positive CAT, the tubal status was subsequently evaluated with hysterosalpingography or laparoscopy, whereas in cases with a negative CAT, tubal pathology was considered absent (Mol *et al.,* 1997). The CAT could be tested with immune-fluorescence technique (IF) or with enzyme immune assays (BioMerieu, Paris, France; Medac GmbH, Wedel, Germany; Savyon Diagnostics Ltd., France). The CAT was considered to be positive for immune-fluorescence technique if the titer was greater than 1:16 and for ELISA if the level was greater than 1.1. Women with two-sided tubal pathology were excluded from the study.

After completion of the fertility work-up, the chance of spontaneous pregnancy within 1 year, leading to live birth, was calculated with a validated prediction model (www.freya.nl/probability.php) (Hunault *et al.,* 2004; Van der Steeg *et al.,* 2007). Couples in whom the probability of spontaneous pregnancy was 40% or greater within 12 months were counseled for expectant management for a period of at least 6 months. After 6 months of expectant management, it was up to the couples to decide to start treatment or to wait for a longer period. Couples with a probability less than 30% were counseled for treatment according to the national fertility guidelines, whereas those with a probability between 30% and 40% were asked to participate in a randomized clinical trial comparing intrauterine insemination with ovarian hyperstimulation with expectant management (Dutch Society of Obstetrics and Gynaecology 2000; Dutch Society of Obstetrics and Gynaecology 1998; Steures *et al.,* 2006).

Follow-up started at completion of the basic fertility work-up and ended after 12 months. The primary end point was spontaneous conception resulting in ongoing pregnancy. The first day of that menstrual cycle was considered to mark the end of time until spontaneous pregnancy. Spontaneous ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal sonography at a gestational age of at least 12 wk, resulting from treatment-independent conception.

Time to pregnancy was considered to be censored when treatment started or at the last date of contact during follow-up and ongoing pregnancy had not occurred. For all couples that were lost to follow-up, the general practitioner was sent a questionnaire on the last known reproductive status of the couple.

Analysis

Female age, menstrual cycle length and basal FSH level are all likely to be related to ovarian aging (Miro *et al.,* 2004). Therefore, we decided to analyze these variables together.

First, we assessed the linearity of these three variables in relation to time to spontaneous ongoing pregnancy using spline functions (Harrell, Jr. *et al.,* 1988). Variables with nonlinear associations were redefined, based on these spline functions.

We then analyzed the predictive value of female age, cycle length, and basal FSH level using Cox proportional hazard modeling of the time to spontaneous ongoing pregnancy. The results were expressed as a hazard ratio (HR). Thereafter, we checked whether the results of the analysis could be confirmed in a multivariable Cox proportional hazard model that included other prognostic factors.

Second, we aimed to estimate the clinical impact if FSH were to be used in a prediction model for pregnancy. For this purpose, it was considered appropriate that couples with a probability of less than 30% on a spontaneous pregnancy were counseled for treatment, whereas couples with a probability of 30% or greater were counseled for expectant management. This policy is supported by the findings of a recently published randomized clinical trial, which showed that above a probability of 30% fertility treatment (intra-uterine insemination with mild ovarian stimulation) did not improve pregnancy rates, compared with expectant management (Steures *et al.,* 2006).

To estimate the added value of basal FSH, we calculated two probabilities of spontaneous pregnancy for each couple. The first probability was calculated with a validated prediction model prediction for pregnancy (model Hunault) (Van der Steeg *et al.,* 2007; Hunault *et al.,* 2004). This model incorporates the variables female age, duration of subfertility, primary or secondary subfertility, referral status and semen analysis. The second probability was calculated with the same model, but also including basal FSH as prognostic factor (FSH model). We next identified the couples for which the probability of less than 30% changed to a probability of 30% or greater (or otherwise) with this new model. In other words, we assessed how many couples would change from treatment to expectant management (or otherwise).

Calculations and analyses were performed with SPSS® 11.0 (SPSS Inc., Chicago, IL) and S-plus® 6.0 (MathSoft Inc., Seattle, WA) programs.

Results

The study profile is shown in figure 1. We registered 7108 subfertile couples in 19 hospitals. Of these couples, 6072 (85%) completed their fertility work-up. In 5380 women (89%), basal FSH was measured. In 288 couples the duration of subfertility was less than 1 year, 636 had an ovulation disorder, 685 had a severe male factor, whereas 251 had double-sided tubal pathology, leaving 3519 couples (58%) that fulfilled the inclusion criteria. Baseline characteristics are presented in table 2. Median basal FSH level was 6.6 IU/L (5^{th} to 95^{th} percentile: 3.5-12). A basal FSH less than 8 IU/L was found in 72% of all women, between 8 and 10 IU/L in 17%, between 10 and 15 IU/L in 8.9%, and 15 IU/L or greater in 2.5% of all women.

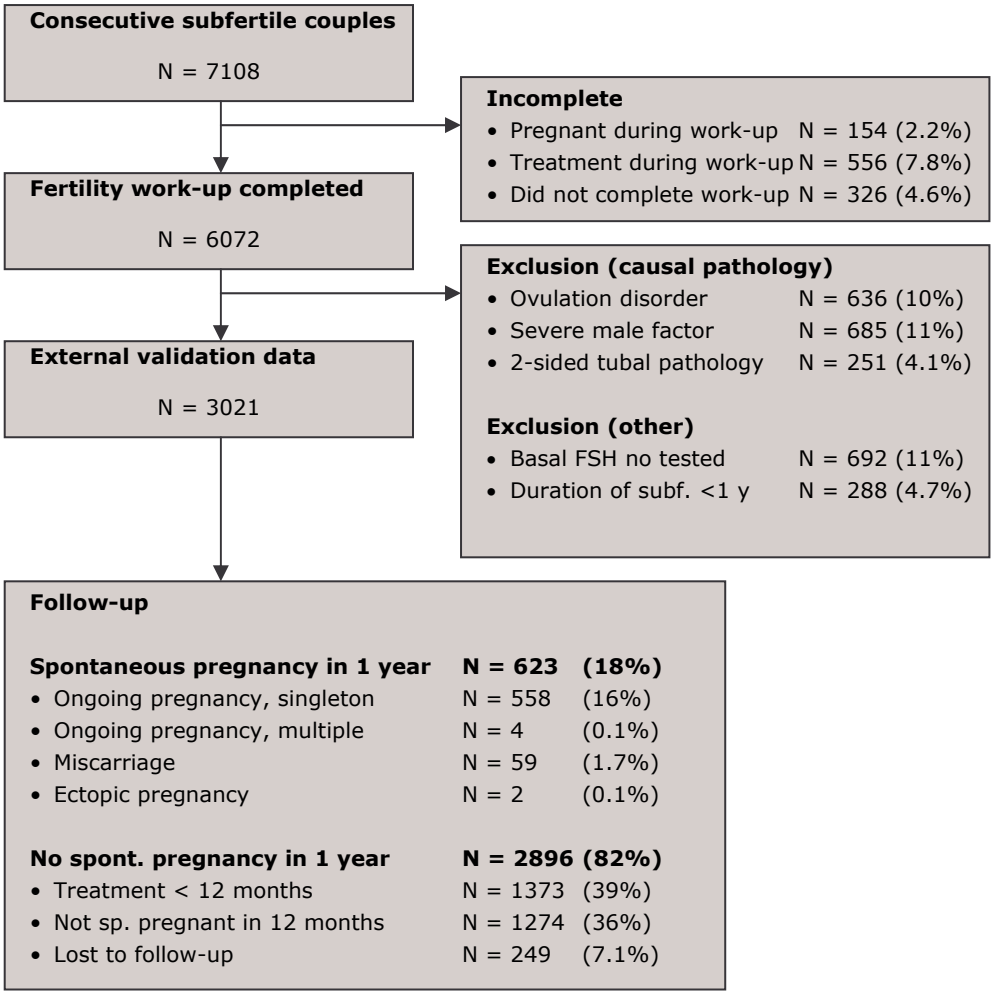
Table 2 Baseline characteristics.

	Mean	**Median**	**5^{th} to 95^{th} percentile**
Female age (y)	32.6		25–40
Male age (y)	35.3		27–45
Duration of subfertility (y)		1.5	1.0–4.1
Subfertility, primary (%)	72		
Semen analysis – TMC (10^6)		54.0	5.6–304
Cycle length (d)		28.1	23–33
FSH (IU/L)		6.6	3.5–12

TMC = total motile sperm count; FSH = follicular stimulating hormone.

The follow-up status of all patients at 12 months is shown in figrue 1. We completed follow-up for 3270 couples (93%). Of all couples, 562 (16%) had a spontaneous ongoing pregnancy within 1 year, including four (0.7%) multiple pregnancies. Unsuccessful pregnancies occurred in 61 couples (1.8%), of which 59 (1.7%) miscarried and two (0.1%) resulted in an ectopic pregnancy, *i.e.* 9.4% and 0.3% of all pregnancies, respectively. Within 12 months 1373 of all couples (39%) started treatment, whereas 1274 (36%) had neither started treatment nor became pregnant. We did not complete follow-up at 12 months for 249 couples (7.1%).

Figure 1 Flow chart.



For female age and cycle length, we found a linear association with the probability of spontaneous ongoing pregnancy. In contrast, a nonlinear association was observed for basal FSH level (figure 2). At less than 8 IU/L, FSH level was not associated with the probability of spontaneous ongoing pregnancy, whereas above 8 IU/L the probability of a spontaneous ongoing pregnancy decreased. We therefore divided basal FSH level into two continuous variables, the first ranging from 0 to 8 and the second starting at a basal FSH level of 8 IU/L.

Female age, cycle length, and FSH concentration, starting at a level of 8 IU/L, were each associated with a decreased probability of spontaneous ongoing pregnancy (table 3). In the multivariable analysis, they remained significant predictors (table 3). At a per-unit increase of FSH over 8 IU/L, the probability of a spontaneous pregnancy decreased with 7%. This translates to a 40% reduction of the probability to conceive in women with a basal FSH of 15 IU/L, and a reduction of 58% in women with a FSH of 20 IU/L, in comparison with women with a basal FSH below 8 IU/L.

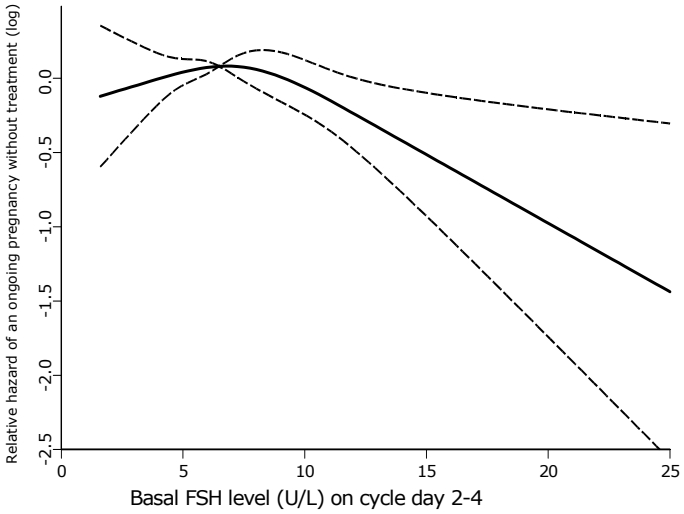


Figure 2 Spline function of basal FSH level. This figure expresses the association between basal FSH levels on cycle day 3 vs the relative hazard (fecundity ratio) of a spontaneous ongoing pregnancy. Dotted lines represent 95% CI.

Table 3 Results of the univariable and multivariable Cox' proportional hazard analysis.

	Univariable analysis		Multivariable analysis	
	HR	**95% CI**	**HR**	**95% CI**
Female age (y)	0.96	(0.94-0.97)	0.97	(0.95-0.99)
Cycle length (per day shorter)	0.95	(0.92-0.99)	0.96	(0.93-1.0)
Basal FSH level until 8 (per IU/L)	1.0	(0.96-1.1)	-	
Basal FSH level above 8 (per IU/L)	0.93	(0.87-0.98)	0.93	(0.87-0.99)

HR = hazard ratio; CI = confidence interval.

To assess possible change in treatment policy, we calculated in 3219 couples a probability for the occurrence of spontaneous pregnancy within 1 year based on the model of Hunault (model Hunault). The probability was compared with the probability calculated with the same model including basal FSH as prognostic factor (FSH model). Probabilities are plotted in a scatterplot (figure 3). With the model of Hunault 1242 couples (39%) had a probability of 30% or greater, whereas the remaining 1977 couples (61%) had a probability less than 30%. When probabilities were calculated with the FSH model (including FSH), 1255 couples (39%) had a probability of 30% or greater, whereas the remaining 1964 couples (61%) had a probability less than 30%. In a total of 97 of the 3219 couples (3.0%) the probability changed to 30% or greater to less than 30% (42 [1.3%]) or the other way around (55 [1.7%]). As a consequence, with a probability of less than 30% as threshold for counseling treatment, 42 couples (1.3%) would be counseled for fertility treatment instead of expectant management, and 55 couples (1.7%) would be counseled for expectant management instead of fertility treatment. So in total in less than 5% of all couples, FSH altered treatment management.

Another approach was the calculation of the fraction of couples in which the probability of a spontaneous pregnancy decreased more than 5% and more 10% due to their FSH level. In 21 of 3219 couples (0.6%) the probability decreased by 10% or more with the prediction model including FSH, whereas in 73 of 3219 couples (2.3%), the probability decreased 5% or more.

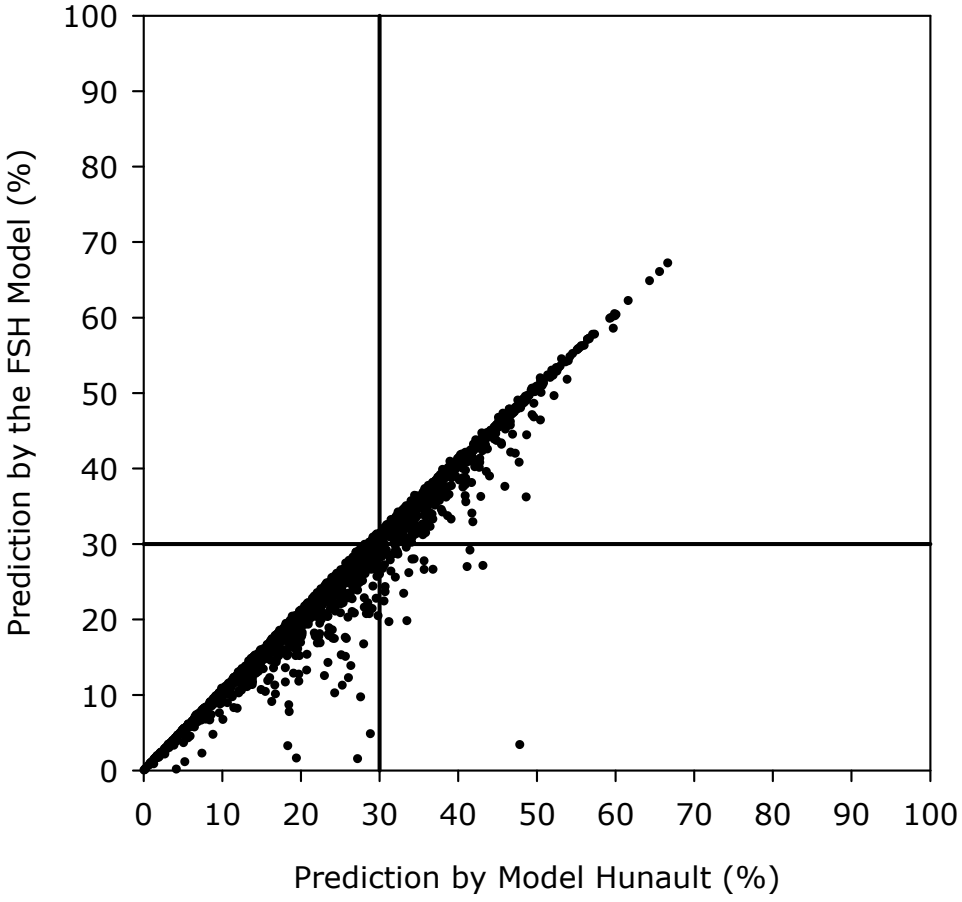


Figure 3 Scatter plot. Correlation between probability based on the model of Hunault and the FSH model. Predictions represent the probabilities of a spontaneous ongoing pregnancy in 12 months. Model Hunault consists of the variables female age, duration of subfertility, previous pregnancy, semen analysis, and referral status (www.freya.nl/probability.php). The straight lines correspond to the threshold of a prediction of 30% below which counseling for fertility treatment is reasonable and vice versa.

Discussion

This study assessed the predictive value and clinical impact of basal FSH in subfertile ovulatory women. Our results show that increased basal FSH levels are associated with a decreased fecundity, even when corrected for cofactors female age and cycle length. Female age, cycle length, and basal FSH were all independent predictors for spontaneous pregnancy. Yet addition of basal FSH to the Hunault prediction model for spontaneous pregnancy did not lead to a large number of clinically relevant changes in pregnancy chances.

Many studies reporting on basal FSH have introduced cutoff values, *i.e.* 10, 15, or 20 IU/L (Van Montfrans *et al.,* 2000; Van Rooij *et al.,* 2004). In contrast, we analyzed basal FSH as a continuous variable. In fact, pregnancy rates in relation to FSH level appear to follow a continuum, with subtle declines noted from as low as 8 IU/L, thus not justifying a single cutoff value basal FSH.

A possible limitation of our study is that we assessed the basal FSH level only once for each woman. A review reported that intercycle variation and hourly variations may result in disparate FSH (Lambalk and De Koning 1998). Maybe a repeat of measurements would alter the accuracy that we report, although it was shown in another study focused on IVF treatment that repeated FSH measurements did not provide a substantial benefit in predicting poor ovarian response in IVF (Bancsi *et al.,* 2004).

Another potential limitation is the fact that different FSH assays were used in the 19 participating clinics. Addition of the type of FSH assay to the Cox regression analysis, however, did not alter the results (data not shown).

A receiver operating characteristic curve analysis was not performed for two reasons; first, receiver operating characteristic analysis does not take time to event, *e.g.* time to spontaneous pregnancy, into account; second, the area under the curve is a measure of discrimination, or the ability to separate two groups, such as case-patients and controls. The sensitivity and specificity have no direct diagnostic meaning for the patient, and the issue is not the risk for having a positive test result but the risk for getting pregnant. The predicted probability, given the risk factors (basal FSH, female age, cycle length) or the posttest probability, can be more useful clinically in assessing future chance to conceive spontaneously than sensitivity or specificity (Cook *et al.,* 2006).

Day 3 estradiol was advocated but not obligatory in our study protocol, but 13 of the 19 participating hospitals assessed day 3 estradiol routinely. Data on 1991 estradiol measurements of the 3519 women (57%) showed estradiol level not to be related to time to spontaneous ongoing pregnancy (HR 0.75; 95% confidence interval (CI) 0.32 – 1.7).

Whereas many data exist on the clinical value of basal FSH in relation to IVF, only two retrospective studies and a small prospective study have reported on basal FSH in relation to spontaneous pregnancy. The first study, performed in a casecontrol setting, failed to demonstrate that basal FSH had predictive capacity (Van Montfrans *et al.,* 2000). In contrast to our study, the primary end point in that study was not specifically spontaneous pregnancy, but pregnancy also including pregnancies after fertility treatment. The same holds true for the only existing prospective study on this subject (Van Rooij *et al.,* 2006). A methodological issue of that prospective study was that they did not perform any regression analysis; they just compared the mean values of basal FSH in the pregnant and non-pregnant group of patients. The second retrospective study reported FSH to be associated with a lower fecundity in women with basal FSH levels above 20 IU/L (Van Rooij *et al.,* 2004), but the study was underpowered to demonstrate an effect over a wider range of FSH values. Because of the large number of subfertile couples in our study, we were able to show a significant decline in fecundity with increasing basal FSH levels.

One reason for subfertility of these patients maybe diminished ovarian reserve (Leach *et al.,* 1997). Our study focused on female age, cycle length, and basal FSH as markers for ovarian aging. Apart from these variables, other tests for ovarian aging have been proposed, such as the clomiphene citrate challenge test, GnRH-agonist stimulation test, measurement of antimullerian hormone and basal inhibin B, and antral follicle count (AFC) (Wolff and Taylor 2004; Klein *et al.,* 1996; Seifer *et al.,* 1999; Seifer *et al.,* 2002; De Vet *et al.,* 2002). Of all these tests, the AFC is the only one that has been found to be superior over basal FSH in the prediction of ovarian response in IVF, although the difference is limited (Hendriks *et al.,* 2005c; Hendriks *et al.,* 2005b; Hendriks *et al.,* 2005a; Bukman and Heineman 2001; Broekmans *et al.,* 2006). Whether the AFC should be part of the basic fertility work-up remains subject of further research.

This study showed that the addition of basal FSH to the Hunault prediction model for spontaneous pregnancy did lead to a change in treatment policy in less than 5% of all couples.

However, the test may still be useful in a more selected group of subfertile women. Such a selection could be based on the prognostic profile

of the couples. The fraction of couples in which treatment altered changed from 3.0% to 8.0% when we focused on only those couples that had probability between 25 and 35% chance to conceive, *i.e.* close to the cutoff probability of 30%, used to decide for treatment decisions. The largest change in management occurred in the group of women aged over 38 years. However, because most of these women are usually advised to start treatment, FSH assessment will, generally speaking, not influence management. In women in the age category between 30 and 38 years management would change in 3%. Cost-effective analysis is needed to decide whether routine FSH assessment in this group should be recommended.

In summary, our results show that basal FSH levels above 8 IU/L or greater have predictive value for spontaneous ongoing pregnancy in a general subfertile population, even after taking female age and cycle length into account. Because the number of couples in whom the FSH level alters management decisions is low, we do not recommend routine testing of basal FSH in subfertile couples.

CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Investigators, participating centers in The Netherlands:
P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis; W.A. Schöls; Amersfoort, Meander Medisch Centrum; M.H. Mochtar; Amsterdam, Academisch Medisch Centrum; H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis; P.G.A. Hompes; Amsterdam, Vrij Universiteit Medisch Centrum; J.M. Burggraaff; Emmen, Scheper Ziekenhuis; G.J.E. Oosterhuis; Enschede, Medisch Spectrum Twente; H. Schouwink; Geldrop, St. Anna Ziekenhuis; P.X.J.M. Bouckaert; Heerlen, Atrium Medisch Centrum; F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis; C.J.C.M. Hamilton; 's Hertogenbosch, Jeroen Bosch Ziekenhuis; M. van Hoven; Hilversum, Tergooiziekenhuizen; J.A. Land; Maastricht, Academisch Ziekenhuis Maastricht; J.A.M. Kremer; Nijmegen, UMC St. Radboud; H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis; F.J.M. Broekmans; Utrecht, UMC Utrecht; C.A.M. Koks; Veldhoven, Máxima Medisch Centrum; P. Bourdrez; Venlo/Venray, Vie Curi Medisch Centrum; W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum, The Netherlands

References

Bancsi LF, Broekmans FJ, Looman CW, Habbema JDF, and Te Velde ER. (2004) Predicting poor ovarian response in IVF: use of repeat basal FSH measurement. *J Reprod Med* 49,187-194.

Bancsi LF, Broekmans FJ, Mol BWJ, Habbema JDF, and Te Velde ER. (2003) 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* 79,1091-1100.

Bongaarts J. (1999) The fertility impact of changes in the timing of childbearing in the developing world. *Popul Stud (Camb)* 53,277-289.

Broekmans FJ, Kwee J, Hendriks DJ, Mol BWJ, and Lambalk CB. (2006) A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update* 12,685-718.

Broekmans FJ, Scheffer GJ, Bancsi LF, Dorland M, Blankenstein MA, and Te Velde ER. (1998) Ovarian reserve tests in infertility practice and normal fertile women. *Maturitas* 30,205-214.

Bukman A and Heineman MJ. (2001) Ovarian reserve testing and the use of prognostic models in patients with subfertility. *Hum Reprod Update* 7,581-590.

Collins JA and Rowe TC. (1989) Age of the female partner is a prognostic factor in prolonged unexplained infertility: a multicenter study. *Fertil Steril* 52,15-20.

Cook NR, Buring JE, and Ridker PM. (2006) The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med* 145,21-29.

De Vet A, Laven JS, De Jong FH, Themmen AP, and Fauser BC. (2002) Antimullerian hormone serum levels: a putative marker for ovarian aging. *Fertil Steril* 77,357-362.

Dutch Society of Obstetrics and Gynaecology. (1998) Guideline-Indications for invitro fertilization (IVF). *NVOG-richlijn nr 9,*1-7.

Dutch Society of Obstetrics and Gynaecology. (2000) Guideline-Intra-uterine insemination (IUI). *NVOG-richtlijn nr 29,*1-8.

Dutch Society of Obstetrics and Gynaecology. (2004) Guideline-Basic fertility work-up. *NVOG-richtlijn nr 1,*1-12.

Evers JL. (2002) Female subfertility. *Lancet* 360,151-159.

Ford D and Nault F. (1996) Changing fertility patterns, 1974 to 1994. *Health Rep* 8,39-46.

Forti G and Krausz C. (1998) Clinical review 100: Evaluation and treatment of the infertile couple. *J Clin Endocrinol Metab* 83,4177-4188.

Harrell FE, Jr., Lee KL, and Pollock BG. (1988) Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 80,1198-1202.

Heck KE, Schoendorf KC, Ventura SJ, and Kiely JL. (1997) Delayed childbearing by education level in the United States, 1969-1994. *Matern Child Health J* 1,81-88.

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

Hendriks DJ, Broekmans FJ, Bancsi LF, Looman CW, De Jong FH, and Te Velde ER. (2005b) Single and repeated GnRH agonist stimulation tests compared with basal markers of ovarian reserve in the prediction of outcome in IVF. *J Assist Reprod Genet* 22,65-73.

Hendriks DJ, Mol BWJ, Bancsi LF, Te Velde ER, and Broekmans FJ. (2005c) 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* 83,291-301.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

Klein NA, Illingworth PJ, Groome NP, McNeilly AS, Battaglia DE, and Soules MR. (1996) Decreased inhibin B secretion is associated with the monotropic FSH rise in older, ovulatory women: a study of serum and follicular fluid levels of dimeric inhibin A and B in spontaneous menstrual cycles. *J Clin Endocrinol Metab* 81,2742-2745.

Lambalk CB and De Koning CH. (1998) Interpretation of elevated FSH in the regular menstrual cycle. *Maturitas* 30,215-220.

Leach RE, Moghissi KS, Randolph JF, Reame NE, Blacker CM, Ginsburg KA, and Diamond MP. (1997) Intensive hormone monitoring in women with unexplained infertility: evidence for subtle abnormalities suggestive of diminished ovarian reserve. *Fertil Steril* 68,413-420.

Martin SP. (2000) Diverging fertility among U.S. women who delay childbearing past age 30. *Demography* 37,523-533.

Miro F, Parker SW, Aspinall LJ, Coley J, Perry PW, and Ellis JE. (2004) Relationship between follicle-stimulating hormone levels at the beginning of the human menstrual cycle, length of the follicular phase and excreted estrogens: the FREEDOM study. *J Clin Endocrinol Metab* 89,3270-3275.

Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Van der Veen F, and Bossuyt PMM. (1997) The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 67,1031-1037.

Noord-Zaadstra BM, Looman CW, Alsbach H, Habbema JDF, Te Velde ER, and Karbaat J. (1991) Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *BMJ* 302,1361-1365.

Scott RT, Jr. and Hofmann GE. (1995) Prognostic assessment of ovarian reserve. *Fertil Steril* 63,1-11.

Scott RT, Leonardi MR, Hofmann GE, Illions EH, Neal GS, and Navot D. (1993) A prospective evaluation of clomiphene citrate challenge test screening of the general infertility population. *Obstet Gynecol* 82,539-544.

Seifer DB, MacLaughlin DT, Christian BP, Feng B, and Shelden RM. (2002) Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. *Fertil Steril* 77,468-471.

Seifer DB, Scott RT, Jr., Bergh PA, Abrogast LK, Friedman CI, Mack CK, and Danforth DR. (1999) Women with declining ovarian reserve may demonstrate a decrease in day 3 serum inhibin B before a rise in day 3 follicle-stimulating hormone. *Fertil Steril* 72,63-65.

Sharara FI, Scott RT, Jr., and Seifer DB. (1998) The detection of diminished ovarian reserve in infertile women. *Am J Obstet Gynecol* 179,804-812.

Steures P, Van der Steeg JW, Hompes PGA, Habbema JDF, Eijkemans MJC, Broekmans FJ, Verhoeve HR, Bossuyt PM, Van der Veen F, and Mol BWJ. (2006) Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 368,216-221.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Van Dessel HJ, Bossuyt PM, Van der Veen F, and Mol BW. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,536-542.

Van Montfrans JM, Hoek A, Van Hooff MH, De Koning CH, Tonch N, and Lambalk CB. (2000) Predictive value of basal follicle-stimulating hormone concentrations in a general subfertility population. *Fertil Steril* 74,97-103.

Van Rooij IA, Broekmans FJ, Hunault CC, Scheffer G, Eijkemans MJC, De Jong FH, Themmen APN, and Te Velde ER. (2006) Use of ovarian reserve tests for the prediction of ongoing pregnancy in couples with unexplained or mild male infertility. *Reprod Biomed Online* 12,182-190.

Van Rooij IA, De Jong E, Broekmans FJ, Looman CW, Habbema JDF, and Te Velde ER. (2004) High follicle-stimulating hormone levels should not necessarily lead to the exclusion of subfertile patients from treatment. *Fertil Steril* 81,1478-1485.

Van Weert JM, Repping S, Van der Steeg JW, Steures P, Van der Veen F, and Mol BWJ. (2005) IUI in male subfertility: are we able to select the proper patients? *Reprod Biomed Online* 11,624-631.

Wolff EF and Taylor HS. (2004) Value of the day 3 follicle-stimulating hormone measurement. *Fertil Steril* 81,1486-1488.

World Health Organization (1999) WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. Cambridge University Press, Cambridge, England.

SUBMITTED

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Peter G.A. Hompes
Jan A. M. Kremer
Loes van der Leeuw-Harmsen
Patrick M.M. Bossuyt
Ben Willem J. Mol
Fulco van der Veen
for the CECERM investigators

chapter (09)

CONTROVERSIAL ROLE OF SPERM COUNT IN MALE FERTILITY

a prospective study

Abstract

Background Despite the dramatic increase in the number of couples seeking subfertility treatment, there is still a surprising confusion over the role of the male partner. Male subfertility is traditionally defined quite arbitrarily according to the World Health Organization (WHO) criteria: semen volume ≤ 2 mL, sperm concentration ≤ 20 x 10^6/mL, progressive sperm motility ≤ 25%, sperm morphology ≤ 15% normal forms, or a total sperm count ≤ 40 x 10^6. Above those values the male's role is considered negligible and below those values the couple's problem is defined as "male factor." These criteria for defining male subfertility are used worldwide, but their ability to distinguish men who are fertile from men who are infertile has never been accurately assessed.

Methods Based on a prospective multicenter cohort study of consecutive patients presenting for subfertility, statistical associations between the male partner's semen analysis results (classified according to the WHO criteria), and the probability of fathering a child without any treatment within a time horizon of one year were calculated. Subsequently, semen quality parameters were redefined and their association with the probability of fathering a child without treatment was re-evaluated.

Results We completed follow-up of 3,129 out of 3,345 couples (94%), of whom 517 (17%) had a spontaneous healthy pregnancy with no treatment. The cumulative pregnancy rate for men with normozoospermia defined by routine WHO criteria did not differ significantly from men with supposedly impaired semen parameters (24% versus 23%; $p = 0.74$). However, despite this lack of predictive power for WHO cut-off criteria, the probability of spontaneous conception was nonetheless closely related to standard semen parameters. We did observe lower probabilities of fathering a child for sperm concentrations below 40 x 10^6/mL, total sperm count below 200 x 10^6, and sperm morphology below 20% normal forms. With a multivariable regression model based on these redefined semen parameters we were able to make a finer differentiation between subfertile men with lower and higher probabilities of fathering a child with a range of 7% to 41%.

Conclusions The current WHO criteria for semen quality are invalid for identifying couples who will conceive spontaneously without medical assistance. Nonetheless, the inexpensive standard semen analysis with redefined and graded parameters (rather than crude cut-offs) has very strong predictive value, obviating the need for more expensive and speculative means of measuring male fertility.

Introduction

Subfertility or involuntary childlessness is one of life's great catastrophes. It occurs in more than one out of ten couples (Gibbons, 1911;Hull *et al.*, 1985;Wang *et al.*, 2003;Gnoth *et al.*, 2003). Although female age and duration of subfertility are the most important factors to affect fertility, reduced semen quality is commonly claimed to be a cause of the subfertility in about half of all these couples (World Health Organization, 1987;De Kretser, 1997;Evers, 2002).

To diagnose male factor subfertility, semen analysis is recommended as the initial test in the fertility work-up (Boston (MA), 2003;National Collaborating Centre for Women's and Children's Health, 2004;Nallella *et al.*, 2006). Worldwide, semen analyses are performed and subsequently interpreted according to the World Health Organization (WHO) criteria (World Health Organization, 1999). The WHO criteria are based on the clinical experience of investigators who studied populations of fertile men. The WHO criteria for male subfertility are a semen volume ≤ 2.0 mL, a sperm concentration of ≤ 20 x 10^6/mL, sperm motility (progressive) ≤ 25%, or a total sperm count ≤ 40 x 10^6 (World Health Organization, 1999). Based on these reference values, male subfertility has been labeled as oligozoospermia (sperm concentration of ≤ 20 x 10^6/mL), asthenozoospermia (sperm motility (progressive) ≤ 25%), teratozoospermia (sperm morphology ≤ 15% normal) or a combination of these categories.

Yet the diagnosis of male subfertility based on semen analysis has been hotly contested over the last half century. In 1951, MacLeod *et al.* proposed the first thresholds for sperm counts (> 20 x 10^6/mL, total sperm count > 100 x 10^6), which allowed clinicians to distinguish fertile from infertile men. Due to the fact that many couples conceived without treatment, in whom the men had sperm counts below these thresholds, other studies attempted to find more appropriate thresholds (MacLeod.J. *et al.*, 1951;MacLeod *et al.*, 1953;Zukerman *et al.*, 1977;David *et al.*, 1979;Kovacs *et al.*, 1982;Emperaire *et al.*, 1982;Hargreave *et al.*, 1983;Schoyswan *et al.*, 1983;Baker *et al.*, 1986;Sokol *et al.*, 1987;Sherins, 1988;Silber, 2000). A review reporting on these studies concluded that standard semen parameters are useful in evaluating the degree of 'male factor' in an infertile couple (Silber, 1989). In view of the widespread use of the WHO criteria and the controversy on what actually constitutes male subfertility, it is surprising that the WHO criteria for male factor subfertility

have never been prospectively validated and don't take into account the female confounding factors.

Several studies have attempted to evaluate the accuracy of the WHO criteria, but they were hampered by small sample size, the failure to take spontaneous pregnancy into account and most of them were either case control settings of fertile and infertile couples, or just a retrospective cohort setting of fertile couples (Ombelet *et al.*, 1997;Bonde *et al.*, 1998;Zinaman *et al.*, 2000;Guzick *et al.*, 2001;Menkveld *et al.*, 2001;Nallella *et al.*, 2006). None were prospective among infertile couples.

The all-important clinical question remains therefore unanswered to this day: can the arbitrary WHO criteria be used in an infertile population to identify the degree to which the male is contributing to the subfertility of the couple and to what extent is the male partner, based on sperm analysis, contributing to the subfertility of the couple? To address this question, we collected data in a prospective cohort study of infertile couples.

Materials and methods

The study was designed as a prospective cohort study, performed in 20 hospitals in the Netherlands. The local ethics committee of each participating center gave Institutional Review Board approval. Between January 2002 and February 2004, we invited consecutive infertile couples that had not been evaluated previously. All couples underwent a basic subfertility work-up according to the guidelines of the Dutch Society of Obstetrics and Gynecology (Dutch Society of Obstetrics and Gynaecology, 2004).

Semen analysis

Semen analysis was performed one to two times, after three days of sexual abstinence. Semen samples were collected at the hospital or at the participant's home by masturbation directly into a 50 mL polyethylene jar. Semen samples were analyzed within one hour after ejaculation. After liquefaction, semen volume was measured in a graded tube with 0.1 mL accuracy. Sperm concentration was counted and motility was assessed in a Makler counting chamber or a Bürger-Türk counting chamber at a magnification of x 200. Sperm morphology was classified according to the 1999 WHO criteria (Dutch Society of Obstetrics and Gynaecology, 2004). The accuracy of sperm concentration, motility and morphology assessment was monitored by employing two separate laboratory technicians. Standards of accuracy were also monitored by the Dutch Foundation for Quality Assessment in Clinical Laboratories (www.skml.nl) three times a year.

The WHO criteria classifies men as oligozoospermic who have a sperm concentration ≤ 20 x 10^6/mL, asthenozoospermic who have a sperm motility (progressive) ≤ 25%, and teratozoospermic who have sperm morphology ≤ 15% normal form. We excluded men with azoospermia, defined as the absence of spermatozoa in the ejaculate, from the analysis because clearly azoospermia can easily be defined as male subfertility.

Work-up of the female partner

Age of the female partner was ascertained as well as other relevant factors in the female exam, e.g. assessment of ovulation, basal FSH level testing on cycle day-3, post-coital test (PCT) and assessment of the Fallopian tubes according to the guidelines of the Dutch Society of Obstetrics and

Gynecology (Dutch Society of Obstetrics and Gynaecology, 2004). Ovulation was assessed by means of a basal body temperature chart, measurement of mid-luteal serum progesterone, or by sonographic monitoring of the cycle. Tubal pathology was assessed by a Chlamydia Antibody Test (CAT), and in case of a positive CAT subsequently evaluated with hysterosalpingography (HSG) or laparoscopy, or directly assessed by HSG or laparoscopy. Couples in whom the female partner was diagnosed with anovulation or two-sided tubal pathology, were excluded from the analysis, since these couples clearly cannot conceive without treatment.

Study design

After completion of the fertility work-up, the theoretical probability of a spontaneous pregnancy within one year, leading to live birth, was calculated based on a prediction model we previously published (www.amc.nl/prognosticmodel) (Hunault *et al.*, 2004;van der Steeg *et al.*, 2007). Depending on that theoretical probability couples were counseled for expectant management, or fertility treatment according to the national fertility guidelines (Dutch Society of Obstetrics and Gynaecology, 1998;Dutch Society of Obstetrics and Gynaecology, 2000). All couples were followed prospectively from the completion of the fertility work-up until spontaneous pregnancy or start of treatment within a time horizon of 12 months. The primary endpoint was time to conception after expectant management resulting in a spontaneous ongoing pregnancy. Ongoing pregnancy was defined as the presence of fetal cardiac activity at transvaginal sonography at a gestational age of at least 12 weeks. Time to spontaneous pregnancy was considered censored whenever treatment started, or at the last date of contact during expectant management. In the event of a miscarriage or an ectopic pregnancy the study and follow-up continued. For all couples lost to follow-up the general practitioner was sent a questionnaire and asked about the last known fertility status of the couple.

Data analysis

For each semen parameter, the median and 5^{th} to 95^{th} percentile range was calculated. In addition, for each semen parameter, the proportion of men below and above the current WHO thresholds was assessed, as well as the proportion of men who were theoretically normozoospermic by WHO standards.

In the first part of the analysis, the estimated cumulative spontaneous ongoing pregnancy rates at one year were calculated with Kaplan-Meier survival analysis for each of the categories according to the 1999 WHO criteria: normozoospermia, oligozoospermia, asthenozoospermia, teratozoospermia, oligoasthenozoospermia, oligoteratozoospermia, astheno-teratozoospermia and oligoasthenoteratozoospermia (World Health Organization, 1999).

Differences in cumulative pregnancy between men with semen quality below the WHO thresholds in comparison to normozoospermic men (as reference category) were calculated, using log rank statistics. All men with semen quality below the WHO thresholds to normozoospermic men were compared and next each of the WHO categories separately to normozoospermic men were compared.

In the second part of the analysis, the association between each semen parameter and the probability of fathering a child were re-evaluated, by means of restricted cubic spline functions and Cox regression analyses (Harrell, Jr. *et al.*, 1988). The spline functions express the probability of fathering a child as a flexible function of the semen parameter. Non-linearity of the spline function was tested with ANOVA statistics.

Based on the spline functions, we redefined the semen parameters and set new semen thresholds where applicable. The redefined semen parameters were then evaluated in univariable and multivariable Cox regression analyses. From this multivariable analysis, we predicted spontaneous pregnancy within one year according to sperm count parameters.

In all analyses a P-value of 0.05 was used as to indicate significance. Calculations were performed with SPSS® (SPSS Inc., Chicago, IL) and S-plus® (MathSoft Inc., Seattle, WA) programs.

Results

Participants

During the study period we registered 4,538 infertile couples that had not been evaluated previously and who had completed their fertility work-up. In 286 men (6.3%) semen analysis was not performed, and 65 men (1.4%) were excluded due to azoospermia. In 638 couples (14%) the female partner was diagnosed with anovulation, and in 204 couples (4.5%) there was double-sided tubal pathology. Thus a total of 1,193 couples were excluded from the study.

Baseline characteristics and pregnancy outcomes

Baseline characteristics of the 3,345 included couples are presented in table 1. Their follow-up status at 12 months is shown in table 2. We completed follow-up for 3,129 couples (94%). 2,779 had no spontaneous pregnancy (83%). 1,307 of those couples (39%) started treatment within 12 months, whereas the 1,256 other couples (38%) did not.

566 couples had a spontaneous pregnancy (17%) within one year, resulting in 511 singleton ongoing pregnancies (16%, e.g. 90% of the spontaneous pregnancies), six multiple ongoing pregnancies (0.2%, e.g. 1.1% of the spontaneous pregnancies) and 49 unsuccessful pregnancies (1.5%, e.g. 8.7% of the spontaneous pregnancies). Of the latter, 46 (1.4%, e.g. 8.1% of the spontaneous pregnancies) were miscarriages and three (0.2%) resulted in an ectopic pregnancy.

Semen characteristics

All men had one semen analysis performed, 1,458 men (44%) a second, with a mean number of samples of 1.4. In 2,996 men all semen parameters were assessed, whereas in 349 men (10%) sperm morphology was not documented.

The semen characteristics are shown in table 1. Median semen volume was 3.0 mL, and 29% of all men had a semen volume below the WHO threshold of 2.0 mL. Median sperm concentration was 40.0 (10^6/mL) and 31% of men had a concentration below the WHO threshold of 20.0 (10^6/mL). Median sperm motility was 33%, median sperm morphology 26% and the median total sperm count was 108 (10^6).

Table 1 Characteristics of the patients (N = 3345).

Male characteristics		**Mean**	**5th-95th pc**
Male age (y)		35.0	27 – 45
Duration of subfertility (y) (Median)		1.5	1.0 – 4.1
Subfertility of the couple, primary (N,%)		2379	71%
Previous pregnancies man in previous partnership (N,%)		372	11%
Referral status (N,%)	Own initiative	209	6.2
	General Practitioner	2763	83
	Gynecologist	280	8.4
	Other specialist	93	2.8
Current smoking (N,%)	yes	895	35
	no	1667	65
	missing	783 (23%)	
Current alcohol use (N,%)	yes	1536	61%
	no	997	39%
	missing	822 (24%)	
Current or past disease (N,%)	None	1962	81
	Diabetes type I	16	0.5
	Cardiovascular disease	28	1.2
	HIV	5	0.2
	Liver pathology	8	0.3
	Cancer in past	18	0.7
	Kidney pathology	7	0.3
	Psychological disease	15	0.6
	Asthma	23	1.0
	Testes	135	5.5
	Orchidopexy	61	2.5
	Testicular surgery	37	1.5
	Inflammation of testes	19	0.8
	Pathology of testes	18	0.7
	Other	192	8.0
	Missing	940 (28%)	
Semen parameters		**Median**	**5th-95thpc**
Volume (mL)		3.0	1.0 – 6.0
Concentration (10^6/mL)		40.0	2.0 – 155
Motility (%) (mean)		33	1 – 70
Morphology (% normal) according to WHO		26	2-56
TC (10^6)		108	5-511
Female characteristics		**Mean**	**5th-95th pc**
Maternal age (years)		32.1	25 – 39
Previous pregnancies woman in previous partnership		352	11%
1-sided tubal pathology		318	13.3

WHO = sperm morphology according to the WHO criteria; TC = total sperm count; pc = percentile.

WHO semen parameters and spontaneous pregnancy rates

38% of the men had a motility below the WHO threshold of 25% progressive motile sperm, 32% of men had a morphology below the WHO threshold of 15% normophorms, and 25% had a total sperm count below the WHO threshold of 40 (10^6). As a result, 41% of the men were classified with normozoospermia according to the WHO criteria. The other men were allocated to one of the categories of poor semen quality: 13% had an isolated asthenozoospermia, 11% had an isolated teratozoospermia, 10% had the combination of oligoasthenoteratozoospermia, 7.6% had the combination of oligoasthenozoospermia, 6.3% had an isolated oligozoospermia, 5.7% had the combination of asthenoteratozoospermia, and 5.1% had the combination of oligoteratozoospermia (figure 1).

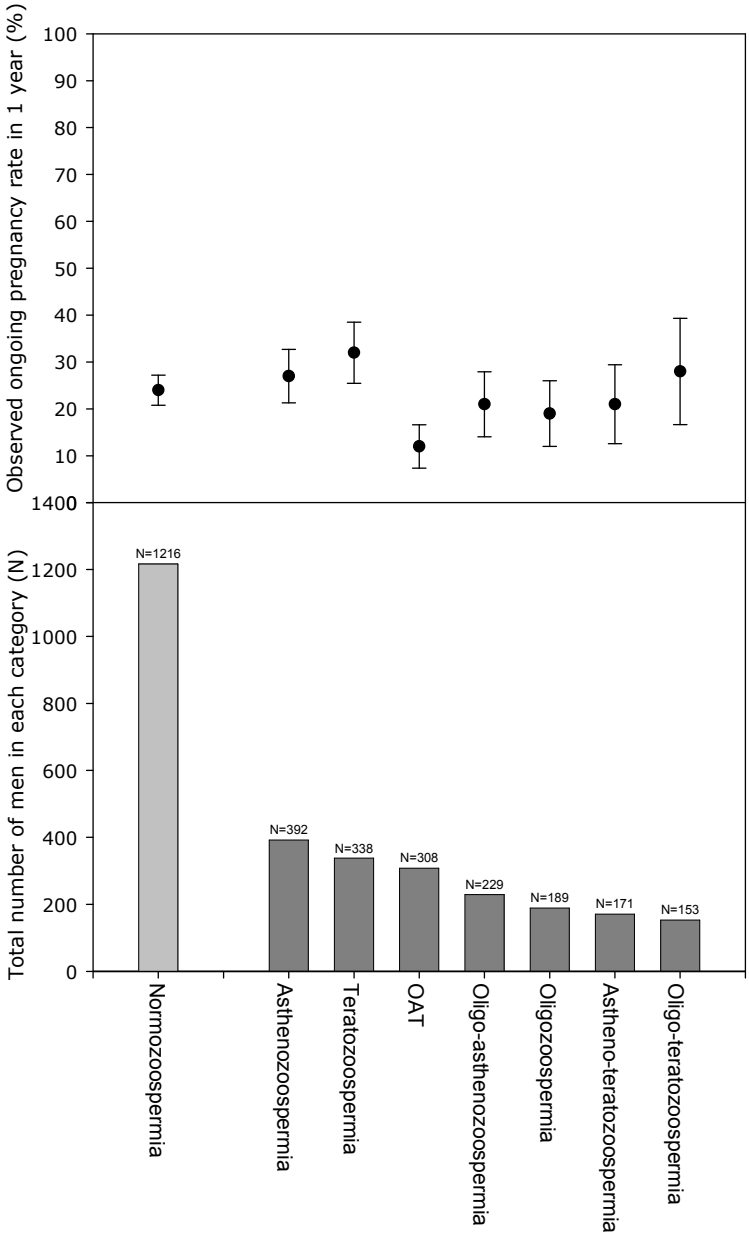
Couples in whom the man had normozoospermia (41%) by WHO criteria had a spontaneous pregnancy rate of 24% (95% CI: 21% to 27%). The cumulative pregnancy rate in the couples (59%) in which the man had sperm quality below the WHO thresholds was 23% (95% CI: 20 to 26) and did not differ significantly (log-rank test $p = 0.74$).

There was no clear pattern between the observed pregnancy rates and the categories of semen impairment. Only oligoasthenoteratozoospermia was associated with a significantly lower cumulative pregnancy rate of 12% (95% CI: 7 to 17) (figure 1).

Table 2 Follow up of the couples (N = 3345).

No spontaneous pregnancy in 1 year	**N (%)**	
Treated within 12 months	1307 (39%)	
Not father a child within 12 months	1256 (38%)	
Lost to follow-up	216 (6.5%)	
Total	**2779 (83%)**	
Spontaneous pregnancy in 1 year	**N (%)**	**Proportion of all spontaneous pregnancies (N = 566)**
Ongoing pregnancy, singleton	511 (15%)	90%
Ongoing pregnancy, multiple	6 (0.2%)	1.1%
Miscarriage	46 (1.4%)	8.1%
Ectopic pregnancy	3 (0.1%)	0.5%
Total	**566 (17%)**	100%

Figure 1 Cumulative spontaneous pregnancy rates at 12 months per subcategory of male subfertility.



Re-evaluation of each semen parameter and spontaneous pregnancy rates

The relation between semen parameters and the probability of spontaneous pregnancy, assessed in the second part of the analysis, is shown in figures 2a to 2e. There was a significantly non-linear association between semen volume and the probability of fathering a child (p = 0.02) (figure 2a). Semen volumes below 2.0 mL were significantly associated with a lower probability of fathering a child. In men with semen volumes above 2.0 mL the probability of fathering a child did not increase any further. The value of 2.0 mL for semen volume was in concordance with the WHO threshold.

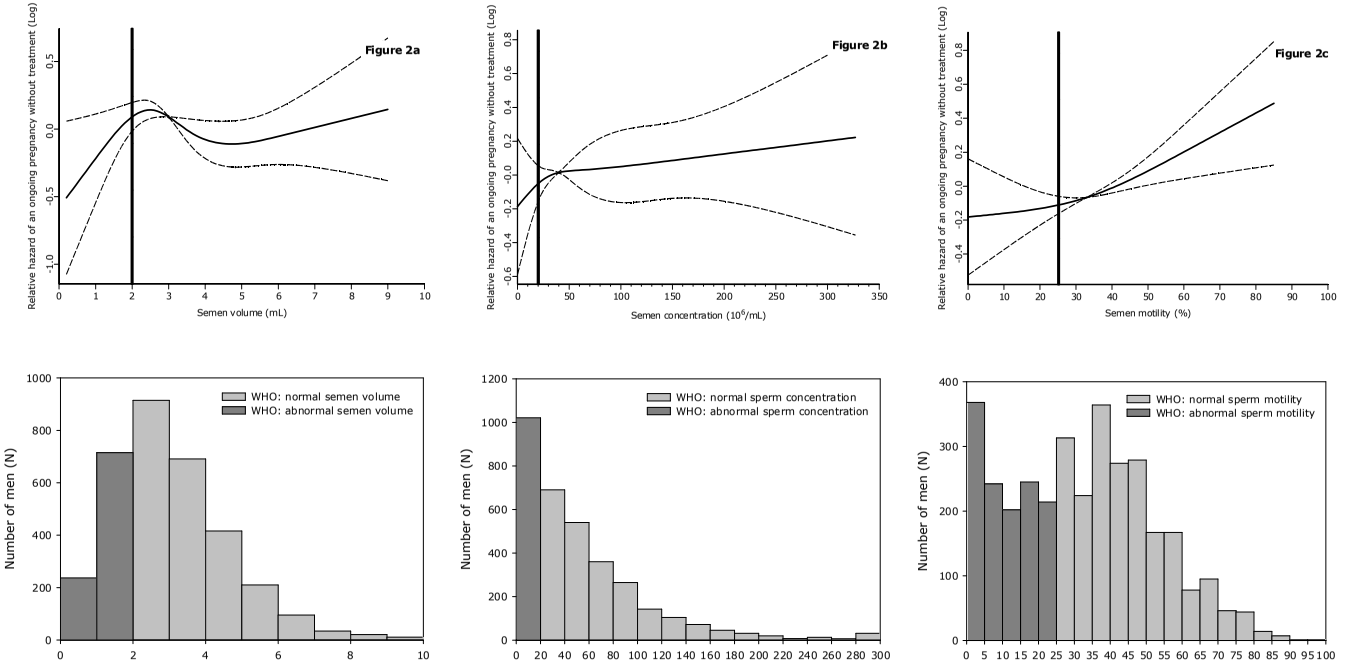
For sperm concentration we also found a non-linear association with the probability of fathering a child (p < 0.01)(figure 2b). However, in contrast to the WHO threshold of 20 x 10^6/mL, we observed a strong relation between sperm concentration and pregnancy rates up to 40 x 10^6/mL, and a stepwise increase in spontaneous pregnancy with increasing sperm concentration in the semen. The threshold was not reached until sperm concentrations over 40 x 10^6/mL wherein the probability of fathering a child then did not increase any further.

Also, sperm motility showed a linear association with the probability of pregnancy with a stepwise increase over the whole range, in contrast to the WHO threshold of 25% progressive motile sperm (figure 2c).

Sperm morphology showed a linear association up to 20% normophorms with a stepwise increase over the whole range, closely to the WHO threshold of 15% (figure 2d).

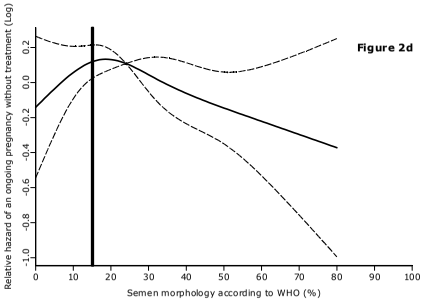
There was a significantly non-linear relation between total sperm count and the probability of fathering a child (p < 0.01) (figure 2e). The pattern of the slope was similar to that of the spline for sperm concentration. Men with low sperm counts, up to 200 x 10^6, had lower probabilities of fathering a child. In men with sperm counts over 200 x 10^6 the probability of fathering a child did not increase any further.

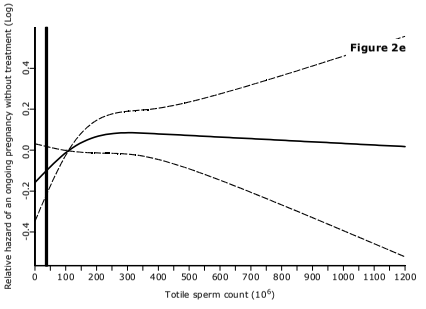
Semen volume, concentration, motility and morphology were each significantly associated with the probability of fathering a child in the multivariable Cox regression analysis (table 3). There was a graded increase in spontaneous pregnancy rate related to the increase in semen parameters, with no discernible cut-off values. Men with lower and higher probabilities of fathering a child ranged from 7% with the poorest sperm quality to 41% for men with the best sperm quality, in contrast to the 23% versus 24% pregnancy rates for men with normal versus abnormal semen based on the WHO criteria (figure 3).

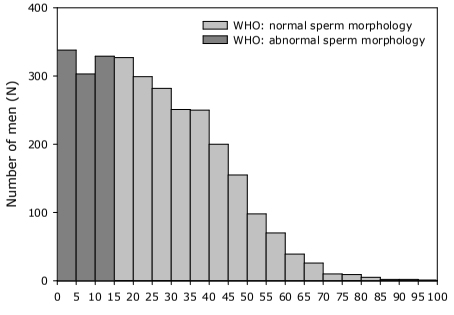


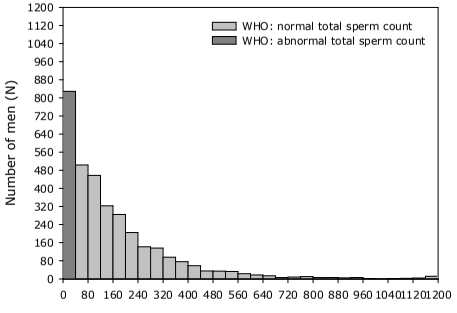
Figures 2a-c Semen parameters and the probability of fathering a child

Spline functions – black vertical lines indicate the reference values for sperm quality as mentioned by the WHO manual 1999; dotted lines represent 95% confidence intervals. Below each figure, the corresponding histograms.









Figures 2d-e Semen parameters and the probability of fathering a child

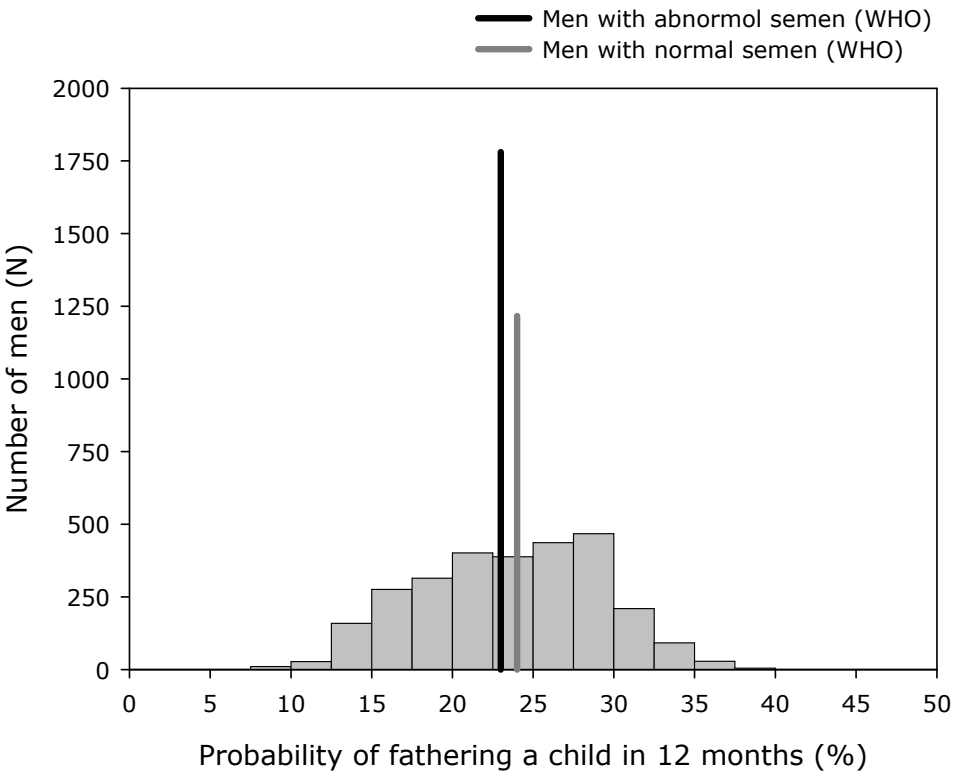
Spline functions – red vertical lines indicate the reference values for sperm quality as mentioned by the WHO manual 1999; dotted lines represent 95% confidence intervals. Below each figure, the corresponding histograms.

Table 3 Pregnancy and male subfertility, based on redefined semen parameters.

			Univariable analysis			Multivariable analysis		
Semen parameters	**N (=3345)**	**%**	**HR**	**95% CI**	***p*-value**	**HR**	**95% CI**	***p*-value**
Semen volume (mL)								
- continuous per mL								
≤ 2.0	952	29	0.78	(0.59-1.02)	0.07	0.70	(0.50-0.98)	0.04
> 2.0	2393	71	1					
Sperm concentration (10^6/mL)								
- continuous per 10 x 10^6 less								
≤ 40	1711	51	0.85	(0.80-0.90)	< 0.01	0.88	(0.77-1.0)	0.05
> 40	1634	49	1					
Sperm motility (%)								
continuous per 10% less	3345	100	0.89	(0.85-0.93)	< 0.01	0.94	(0.89-0.99)	0.01
Sperm morphology (% normal)								
- continuous per 10% less (N = 2996)								
≤ 20	1297	43	0.66	(0.55-0.79)	< 0.01	0.78	(0.64-0.95)	0.01
> 20	1699	57	1					
Total sperm count (10^6)								
- continuous per 10 x 10^6 less								
≤ 200	2406	72	0.97	(0.96-0.98)	< 0.01	-[a]		
> 200	939	28	1			-[a]		

[a] variable not analyzed in the multivariable analysis. HR = hazard ratio; CI = confidence interval.

Figure 3 Histogram, which shows the calculated probabilities of fathering a child with a pregnancy prediction model based on the refined semen parameters. The black line represents the cumulative pregnancy rate at 12 months in men with abnormal semen according to the WHO criteria, the dark gray line represents the cumulative pregnancy rate in men with normal semen.



Discussion

This is the first large prospective study of the relation between semen quality and the probability to father a child spontaneously in infertile couples. We observed that the WHO criteria were inadequate to predict conception, and this is probably a reason there is so much controversy in the field of male subfertility. The cumulative pregnancy rate at one year did not differ between normozoospermic and non-normozoospermic men, which explains why so many clinicians have been puzzled over the role of the sperm count in the couples subfertility. We also demonstrated that semen quality does matter, because our multivariable model, based on redefined semen parameters, was able to discriminate between men with lower and higher probabilities of fathering a child.

Previous studies on the significance of sperm parameters considered semen analysis according to the WHO criteria 'nothing more than an ancestral heirloom' (Menkveld *et al.*, 2001) or stated - without providing the evidence - that 'the WHO criteria failed to meet rigorous clinical, technical, and statistical standards' (Guzick *et al.*, 2001) or simply reported that serious concerns existed about the true value of the WHO semen analysis results and cut-off values (Ombelet *et al.*, 1997;Bonde *et al.*, 1998;Nallella *et al.*, 2006). These arguments led them to establish their own cut-off values for the standard semen parameters. In doing this, they conducted small case control studies of infertile and fertile couples, or cohort studies of fertile couples, overlooking the fact that these cut-off values cannot be translated to couples who have been infertile, trying for a child unsuccessfully for a year. This in an important flaw as predicting fatherhood is of clinical importance only in this group.

Our study cohort is probably representative for subfertility cohorts elsewhere as 59% of all men had abnormal semen parameters according to the WHO criteria, which is in line with the observation that male factor subfertility commonly occurs in about half of all infertile couples (World Health Organization, 1999;De Kretser, 1997). A potential limitation of our study might have been informative censuring during follow-up. It is not unlikely that women of men with poor semen quality were treated with IVF or IUI sooner than women of men with normal semen quality. In that case the proportional hazards assumption would not have been met, leading to biased results of the Cox analysis. However, we found no signs of influential informative censuring in our analysis (data not shown).

In this study we did not correct for female age or female factors in order to make a straight comparison to the WHO criteria, in which these factors are also not taken into account. This may potentially have biased our results as one might speculate that older women may have more often men with lower sperm counts. However, we observed an inversed association between female age and semen quality. All semen parameters, except for semen volume, showed higher values the older the women became (data not shown). For example, the median total sperm count was 80×10^6 in the female age group between 18 and 25 years, whereas it was 116×10^6 in the group between 30 and 35, and was 142×10^6 in the group over 40 years of age. This may be explained by the fact that female age is strongly related to fertility: men with low to normal semen parameters and young female partners will conceive quickly, while men with normal to high semen parameters and female partners around 40 years of age will not.

The fact that the WHO criteria do not correlate well with the probability of conception explains the confusion in the literature and in most centers on what really is male subfertility. In fact, it has led to a plethora of useless commercial innovations like the Hamster test, the sperm penetration tests, and now the DNA fragmentation tests.

The WHO criteria are used in fertility centers all over the world in the decision for invasive and costly fertility treatment, but have never been validated in prospective studies. This important omission has been acknowledged by the WHO as well, who issued a call for large prospective studies on this subject almost a decade ago.

It is clear from this study that in a large infertile population standard norms of semen quality based on crude cut offs of normal versus abnormal are of no use in identifying which couples are likely to conceive without medical assistance and which are not. However, a nihilistic conclusion about the predictive value of the simple sperm analysis is also not warranted, as there is a strong linear correlation between increased sperm parameters and an increased probability of spontaneous conception.

Acknowledgements

The authors thank all participating hospitals for their contribution to this study. Furthermore, Dr. S. Repping, PhD., clinical embryologist (Center for Reproductive Medicine, Academic Medical Center, Amsterdam, The Netherlands), is acknowledged for his critical reading and valuable suggestions on previous drafts of this paper.

CECERM study group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Investigators, participating centers in The Netherlands:
P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis; M.H. Mochtar; Amsterdam, Academisch Medisch Centrum; G.L.M. Lips; Amsterdam, BovenIJ Ziekenhuis; H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis; H.E. Bobeck; Beverwijk, Rode Kruis Ziekenhuis; J.P. Dörr; Den Haag, Westeinde Ziekenhuis; P.J.Q. van der Linden; Deventer, Ziekenhuis Deventer; J.M. Burggraaff; Emmen, Scheper Ziekenhuis; M.H. Schouwink; Geldrop, St. Anna Ziekenhuis; F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis; C.J.C.M. Hamilton; 's Hertogenbosch, Jeroen Bosch Ziekenhuis; M. van Hoven; Hilversum, Tergooiziekenhuizen; M.H. Emanuel; Hoofddorp, Spaarne Ziekenhuis; J.H. Schagen-Van Leeuwen; Nieuwegein, St. Antonius Ziekenhuis; J.A.M. Kremer; Nijmegen, UMC St. Radboud; C. van Katwijk; Purmerend, Waterland Ziekenhuis; H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis; C.A.M. Koks; Veldhoven, Máxima Medisch Centrum; P. Bourdrez; Venlo/Venray, Vie Curi Medisch Centrum; W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum.

References

Baker HWG, Burger HG, and De Kretser D (1986) Male Infertility. In Steinberger E, Frajese C, and Steinberger A (eds) Reproductive Medicine. Raven Press, New York, pp. 187-197.

Bonde JP, Ernst E, Jensen TK, Hjollund NH, Kolstad H, Henriksen TB, Scheike T, Giwercman A, Olsen J, and Skakkebaek NE. (1998) Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet* 352,1172-1177.

Boston (MA). (2003) Infertility. A guide to evaluation, treatment, and counseling. www.guideline.gov.

David G, Jouannet P, Martin-Boyce A, Spira A, and Schwartz D. Sperm counts in fertile and infertile men. *Fertil Steril* 1979;31:453-455.

De Kretser DM. (1997) Male infertility. *Lancet*;349:787-790.

Dutch Society of Obstetrics and Gynaecology. Guideline - Indications for invitro fertilization (IVF). (1998) *NVOG-richlijn nr. 9,*1-7.

Dutch Society of Obstetrics and Gynaecology. Guideline - Intra-uterine insemination (IUI). (2000) *NVOG-richtlijn nr. 29,*1-8.

Dutch Society of Obstetrics and Gynaecology. Guideline - Basic fertility work-up. (2004) *NVOG-richtlijn nr. 1,*1-12.

Emperaire JC, Gauzere-Soumireu E, and Audebert AJ. (1982) Female fertility and donor insemination. *Fertil Steril* 37,90-93.

Evers JL. (2002) Female subfertility. *Lancet* 360,151-159.

Gibbons RA. (1911) A lecture on sterility, its etiology and treatment. *London: J&A Churchill*.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, and Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod* 18,1959-1966.

Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima ST, Coutifaris C, Carson SA, Cisneros P, Steinkampf MP, Hill JA et al. (2001) Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med* 345,1388-1393.

Hargreave TB and Elton RA. (1983) Is conventional sperm analysis of any use? *Br J Urol* 55,774-779.

Harrell FE, Jr., Lee KL, and Pollock BG. (1988) Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 80,1198-1202.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, and Desai KM. (1985) Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 291,1693-1697.

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

Kovacs AT, Lecton JF, Matthews CD, Steigrad SJ, Saunders DM, Jones WR, Lynebain R, and McMaster R. (1982) The outcome of artificial donor insemination compared to the husband's fertility status. *Clin Reprod Fertil* 1,295.

MacLeod J and Gold. (1953) The male factor in fertility and infertility. VI. Semen quality and other factors in relation to ease of conception. *Fertil Steril* 4,10.

MacLeod.J. and Gold RZ. (1951) The male factor in fertility and infertility. II. Sperm counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J Urol* 66,436.

Menkveld R, Wong WY, Lombard CJ, Wetzels AM, Thomas CM, Merkus HM, and Steegers-Theunissen RP. (2001)Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of in-vivo thresholds. *Hum Reprod* 16,1165-1171.

Nallella KP, Sharma RK, Aziz N, and Agarwal A. (2006) Significance of sperm characteristics in the evaluation of male infertility. *Fertil Steril* 85,629-634.

National Collaborating Centre for Women's and Children's Health cbN. Fertility guideline: assessment and treatment for people with fertility problems. www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline 2004.

Ombelet W, Bosmans E, Janssen M, Cox A, Vlasselaer J, Gyselaers W, Vandeput H, Gielen J, Pollet H,

Maes M et al. (1997) Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing. *Hum Reprod* 12,987-993.

Schoyswan R and Gerris. (1983) Twelve year followup: study of pregnancy rates in 1921 couples with idiopathically impaired male fertility. *Acta Eur Fertil* 14,51-55.

Sherins R. (1988) A low sperm concentration does not preclude fertility in men with isolated hypo-gonadotropic hypogonadism after gonadotropin therapy. *Fertil Steril* 50,343-347.

Silber SJ. The relationship of abnormal semen parameters to male fertility. *Hum Reprod* 1989;4:947-953.

Silber SJ. Evaluation and treatment of male infertility. (2000) *Clin Obstet Gynecol* 43:854-888.

Sokol RZ and Sparkes R. Demonstrated paternity in spite of severe idiopathic oligospermia. (1987) *Fertil Steril* 47,356-358.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ, Bossuyt PM, Van der Veen F, and Mol BW. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,536-542.

Wang X, Chen C, Wang L, Chen D, Guang W, and French J. (2003) Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79,577-584.

World Health Organization. Towards more objectivity in diagnosis and management of male infertility. (1987) *Int J Androl* 7,1-53.

World Health Organization (1999) WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. 4th edn, Cambridge University Press, Cambridge, England.

Zinaman MJ, Brown CC, Selevan SG, and Clegg ED. Semen quality and human fertility: a prospective study with healthy couples. (2000) *J Androl* 21,145-153.

Zukerman Z, Rodriguez-Rigau LJ, Smith KD, and Steinberger E. (1977) Frequency distribution of sperm counts in fertile and infertile males. *Fertil Steril* 28,1310-1313.

HUMAN REPRODUCTION
2004; 19: 1373 - 79

Jan Willem van der Steeg
Pieternel Steures
Marinus J.C. Eijkemans
J. Dik F. Habbema
Fulco van der Veen
Patrick M.M. Bossuyt
Peter G.A. Hompes
Ben Willem J. Mol

chapter (10)

SHOULD THE POST-COITAL TEST (PCT) be part of the routine fertility work-up?

Abstract

Background This study aimed to determine whether medical history and semen analysis can predict the result of the post-coital test (PCT).
Methods A previously reported data set of Dutch patients collected between 1985 and 1993 was used. Our study was limited to just patients with an ovulatory cycle. Data were complete for medical history, semen analysis and PCT. We performed logistic regression analysis to evaluate whether these factors could predict the result of the PCT (PCT model). Furthermore, we evaluated the additional contribution of the PCT in the prediction of treatment-independent pregnancy (pregnancy model).
Results Thirty-four percent (179 out of 522) had an abnormal PCT. The PCT model contained previous pregnancy (odds ratio (OR): 2.1; 95% confidence interval (CI): 1.3-3.5), semen volume (OR: 0.88; 95% CI: 0.77-0.99), sperm concentration (OR: 0.96; 95% CI: 0.94-0.97), sperm motility (OR: 0.97; 95% CI: 0.96-0.98) and sperm morphology (OR: 2.7; 95% CI: 1.2-6.8). The area under the ROC curve of the model was 0.81. In the pregnancy model, the result of the actual PCT could be replaced by the predicted result of the PCT model in about half of the couples, without compromising its predictive capacity.
Conclusions The medical history and semen analysis can predict the result of the PCT in ~ 50% of the subfertile couples with a regular cycle, without compromising its potential to predict pregnancy.

Introduction

The use of the post-coital test (PCT) in the basic fertility work-up has been subject to debate over the last ten years (Cohlen *et al.,* 1999; Oei 1998; Eimers *et al.,* 1994; Griffith and Grimes 1990). The PCT is traditionally used as a test to diagnose cervical factor subfertility. An abnormal PCT in the presence of normal semen is considered to reflect cervical hostility.

In the absence of tubal occlusion, some argue that cervical factor subfertility can be treated with intrauterine insemination (IUI). Among the five randomized studies in which IUI is compared with no treatment, two clearly reported a beneficial result of IUI (Check *et al.,* 1995; Te Velde *et al.,* 1989; Check *et al.,* 1995; Te Velde *et al.,* 1989), whereas three other studies indicated the absence of such a beneficial result (Kirby *et al.,* 1991; Friedman *et al.,* 1989; Chaffkin *et al.,* 1991). If IUI were to be effective, the PCT has potential value in the work-up for subfertility. However, the routine use of the PCT has been reported to lead to more treatment without an associated increase in pregnancy rates (Oei *et al.,* 2001). In this view, the routine use of the PCT in the basic fertility work-up would be unnecessary.

Apart from the diagnostic aspect of the PCT, some authors have also emphasized the prognostic value of the PCT in the prediction of treatment-independent pregnancy. An abnormal PCT would decrease the probability of treatment-independent pregnancy 2- to 3-fold (Eimers *et al.,* 1994; Oei *et al.,* 1995a; Snick *et al.,* 1997). However, despite these findings, Eimers *et al*. advocated the use of the PCT in the routine fertility work-up as a prognostic test, whereas Oei *et al*. argued against the routine performance of the PCT (Oei *et al.,* 1996). Only one author made a distinction between couples with a long duration of subfertility and couples with a shorter duration of subfertility (Glazener *et al.,* 2000). It was concluded that the PCT has additional prognostic value, but only in couples with a duration of subfertility of less than three years.

Therefore, there is at present no clear-cut evidence that the PCT has diagnostic or prognostic value as a routine test for subfertile patients. However, the PCT might be of use for some couples, whereas it is not useful for other couples. Therefore, if one were able to select a group of subfertile couples, based on their history and semen analysis, in which the PCT has limited additional value, the routine performance of this test would not be necessary. In the remaining couples, the routine performance of this test could be evaluated by means of a randomized controlled trial.

The present study aims to evaluate whether the results of the PCT can be predicted from the patients' history and the semen analysis. Furthermore, we assessed whether omission of the PCT in couples in whom the PCT result can be predicted would compromise the capacity to predict treatment-independent pregnancy from data obtained at the basic fertility work-up.

Materials and methods

We used the data reported on previously by Snick *et al.* (Snick *et al.,* 1997). This data set contains information on consecutive couples referred by general practitioners, presenting with subfertility between January 1, 1985 and December 31, 1993 to the Walcheren hospital, Vlissingen, The Netherlands. Subfertility was defined as failed conception after one year of unprotected intercourse.

Since the PCT cannot be performed in anovulatory women, we limited data to patients with an ovulatory cycle. The menstrual cycle was considered ovulatory if the duration of the menstrual cycle was < 8 weeks and if serum progesterone was above 18 nmol/l in the luteal phase. Data had to be complete for medical history and semen analysis. The medical history consisted of data on male and female age, subfertility being primary or secondary, and duration of subfertility. Secondary subfertility was defined as subfertility in a couple that had a pregnancy in the past. Duration of subfertility was defined as the time since a couple had unprotected intercourse till the time they visited the hospital. A semen analysis and PCT were performed in each couple. From the semen analysis, data were available on the volume of the sample, the concentration of sperm cells (10^6/mL), the percentage of progressively motile sperm cells and the morphology (categorized in two groups; < 20% or > 20%). The exact PCT procedure has been described previously by Snick *et al.* (Snick *et al.,* 1997). In brief, the PCT was scheduled 15 or 16 days before the expected onset of the next menstrual period, and repeated every 48 h until the test was normal, or until either the basal body temperature showed a persistent rise, or the dominant follicle had disappeared on ultrasound. A PCT defect (abnormal PCT) was diagnosed if the best test showed not more than one forward-moving spermatozoon in the whole cervical mucus sample, 8 to 16 hours after intercourse. All other outcomes of the PCT were considered to be normal.

All couples were followed until the occurrence of pregnancy. Pregnancy was defined as amenorrhoea of ≥ 6 weeks in combination with a positive urine pregnancy test, or embryonic sac visible at sonography. Time to pregnancy was considered to be censored when treatment was started, or at the last date of contact during follow-up, when the patient was not pregnant.

Analysis

Prediction of the PCT

To evaluate whether the results of the PCT could be predicted by data from the medical history and the semen analysis, we studied the associations between these variables and the result of the PCT. First, we assessed the linearity of the continuous variables duration of subfertility, male and female age, the semen parameters volume, concentration, and motility on one hand, and the result of the PCT on the other hand, using spline functions (Harrell, Jr. *et al.,* 1988). Non-linear associations were redefined, based on these spline functions. We then performed univariable and multivariable logistic regression analysis with the result of the PCT as dependent variable. For the multivariable regression analysis we used a stepwise backwards selection procedure. Usually, the selection of variables is performed with a significance level of 5%. As the incorrect exclusion of a factor would be more deleterious than including too many factors, we selected all prognostic variables with a significance level up to 30% (Steyerberg *et al.,* 1999).

To evaluate the performance of the logistic model, the area under the receiver operating characteristic (ROC) curve was calculated. Sensitivity was defined as the fraction of patients with an abnormal PCT that was predicted correctly, whereas specificity was defined as the fraction of patients with a normal PCT that was predicted correctly.

Internal validation was performed with bootstrapping. Bootstrapping is a technique to create new data sets by random drawing from the sample with replacement. In each of these new data sets (N = 1000) the same multivariable regression was assessed. By analysing the difference of the prognostic models a shrinkage factor was calculated to reduce the overfit of the created model. Calibration was evaluated with the Hosmer and Lemeshow goodness-of-fit test statistic.

Prediction of pregnancy

To evaluate whether the PCT, performed in a selected group, can predict treatment-independent pregnancy, we used Cox proportional hazard analysis to develop three pregnancy prediction models. The initial model contained the variables duration of subfertility, age of the female partner, previous pregnancies for the couple and the semen parameters volume, concentration, motility and morphology, but not the PCT (pregnancy model I). We then developed a second model, in which both semen analysis and PCT were always performed (pregnancy model II). Finally, we developed a third pregnancy prediction model (pregnancy model III), in

which the PCT was only performed in cases where the probability of an abnormal PCT was in between the cut-off points 0.2 and 0.6. These points correspond to the chance of an abnormal PCT of 20% and 60%, calculated with the PCT model. In cases where the chance was < 20% or > 60%, pregnancy model III used the prediction for treatment-independent pregnancy made without the PCT.

The performances of the models I, II and III were assessed by comparing the area under the curve (AUC) of the three ROCs. The AUCs were calculated, using the Kaplan-Meier method as proposed by Heagerty *et al*. (Heagerty *et al.,* 2000), and the 95% confidence intervals (CI) were calculated by bootstrapping.

Results

The data set contained 726 patients of whom 587 (81%) couples had an ovulatory cycle. The PCT was available in 531 couples (90%). In nine couples data were incomplete. The baseline characteristics of the 522 couples that were included in the study are shown in the table.

There were 179 (34%) couples with an abnormal PCT. Female age, male age as well as duration of subfertility were almost equal for the subgroup with a normal PCT and the subgroup with an abnormal PCT. After the PCT was performed, tubal pathology was diagnosed in 109 (22%) women, in whom 55 (11%) had reconstructive surgery of the Fallopian tubes. During the follow-up period 228 spontaneous pregnancies were registered, comprising 31 spontaneous abortions, seven ectopic pregnancies and 190 ongoing pregnancies.

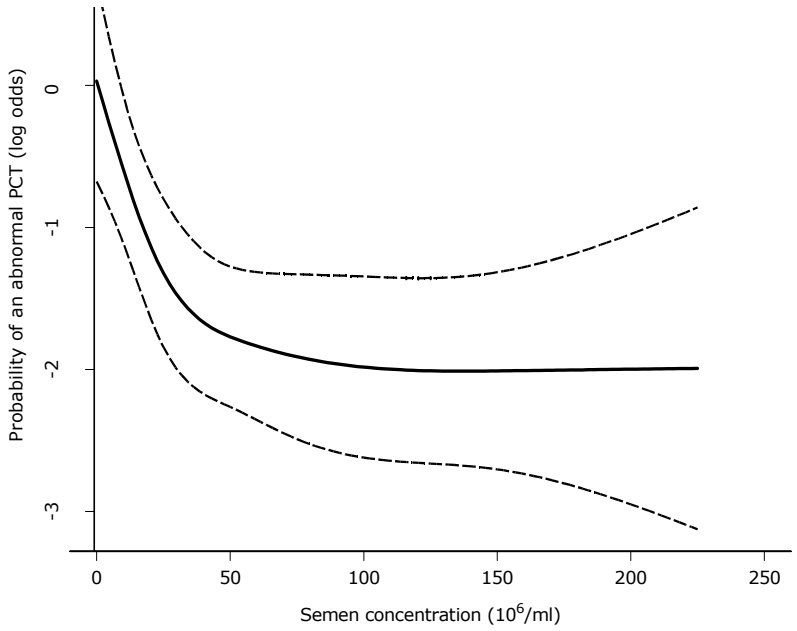


Figure 1 Spline function, which visualizes the association of the continuous variable sperm concentration and the probability of an abnormal PCT.

Table Baseline characteristics of the Snick data – couples with an ovulatory cycle, and results of the univariable and multivariable regression analysis.

	Result of the PCT				Univariable analysis		Multivariable analysis	
	Abnormal 179 (34%)		Normal 343 (66%)		OR	95% CI	OR	95% CI
History								
Female age (mean, range)	28.8	(18-40)	29.9	(20-42)	0.95	(0.91-0.99)		
Male age (mean, range)	30.7	(22-60)	31.7	(21-50)	1.0	(0.96-1.04)		
Duration of subfertility (y) (median, range)	1.5	(0.9-12)	1.3	(0.8-12)	1.0	(0.89-1.2)		
Primary subfertility	149	(83%)	221	(64%)	2.7	(1.7-4.2)	2.1	(1.3-3.5)
Semen characteristics								
Volume (mL) (median, range)	3	(0-8)	3	(0-12)	0.85	(0.76-0.95)	0.88	(0.77-0.99)
Concentration (10^6/mL) (median, range)	12	(0-300)	47	(0-400)	0.94[a]	(0.92-0.95)	0.96	(0.94-0.97)
Motility (%) (median, range)	15	(0-80)	40	(0-85)	0.96	(0.95-0.97)	0.97	(0.96-0.98)
Abnormal morphology (< 20%)	25	(14%)	10	(3%)	4.9	(2.3-10.6)	2.7	(1.2-6.8)

OR = odds ratio; CI = confidence interval.

[a] OR for sperm concentrations from 0 to 40 x 10^6/mL. Higher concentrations have equal OR as concentrations of 40x10^6/mL.

The chance of an abnormal PCT can be calculated from the multivariable model with the formula:

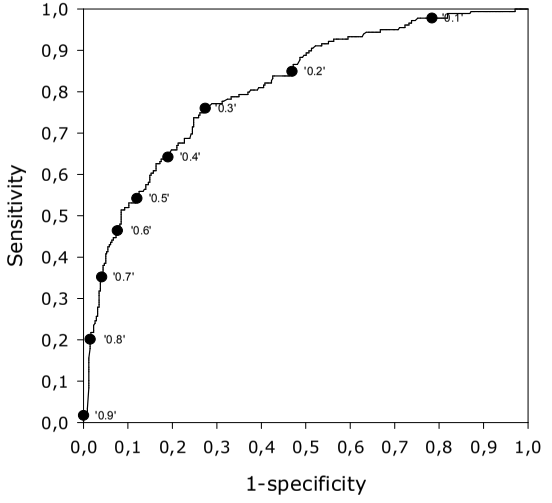
Probability = $1/(1+\exp(-\beta))$, where

β = 1.048 + primary subfertility x 0.74 + semen volume x -0.13 + sperm concentration x -0.041 + motility x -0.03 + abnormal morphology x 0.99.

For female and male age, duration of subfertility, semen volume and sperm motility, we found a linear association with the result of the PCT on the logistic scale. In contrast, a non-linear association was observed for the sperm concentration and the result of the PCT on the logistic scale (figure 1). We therefore decided not to distinguish between sperm concentrations over 40 x 10^6/mL, and to model lower values in a linear way.

The table also shows the results of the univariable and multivariable analyses. Primary subfertility of the couple was associated with an increased probability of an abnormal PCT as were low semen volume, lower semen concentrations, low sperm motility and abnormal sperm morphology. These factors were therefore included in the PCT prediction model. Internal validation by bootstrapping showed a shrinkage factor of 0.91, which corresponds to 9% overfit of the initial model. All odds ratios (OR) and the intercept of the multivariable model were adjusted for this 9% overfit.

The PCT prediction model had an area under the ROC curve of 0.81 (95% CI: 0.77-0.85), indicating an excellent discriminative performance (figure 2a). Figure 2b shows how the number of normal and abnormal PCTs depends on the probability of an abnormal PCT (divided into 10 groups of 10% prediction chance). This figure shows that 27 of the 209 patients (13%) in whom the calculated probability of an abnormal PCT was under 20% indeed had an abnormal PCT. On the other hand, the PCT result was abnormal in 83 of the 109 patient (76%) in whom the calculated probability of an abnormal PCT was 60% or higher. Tubal pathology was diagnosed in 59 couples that had a calculated probability of an abnormal PCT result of < 20%, in 37 couples with a calculated probability in between 20 and 60%, and in 13 couples with a calculated probability over 60%.



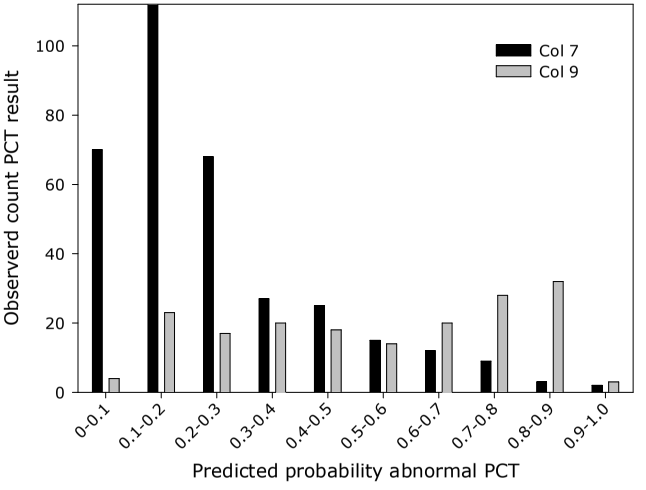


Figure 2 (**a**) Receiver operating characteristic (ROC) curve of the multivariable logistic regression model for the prediction of an abnormal PCT. Points located on the curve correspond to the cut-off values of the predicted PCT (e.g. a cut-off value of 0.1 considers patients with a probability below 0.1 to have a normal PCT, and above 0.1 to have an abnormal PCT). (**b**) Clustered bar chart. For each probability of an abnormal PCT the distribution of total observed normal and abnormal PCT's is shown.

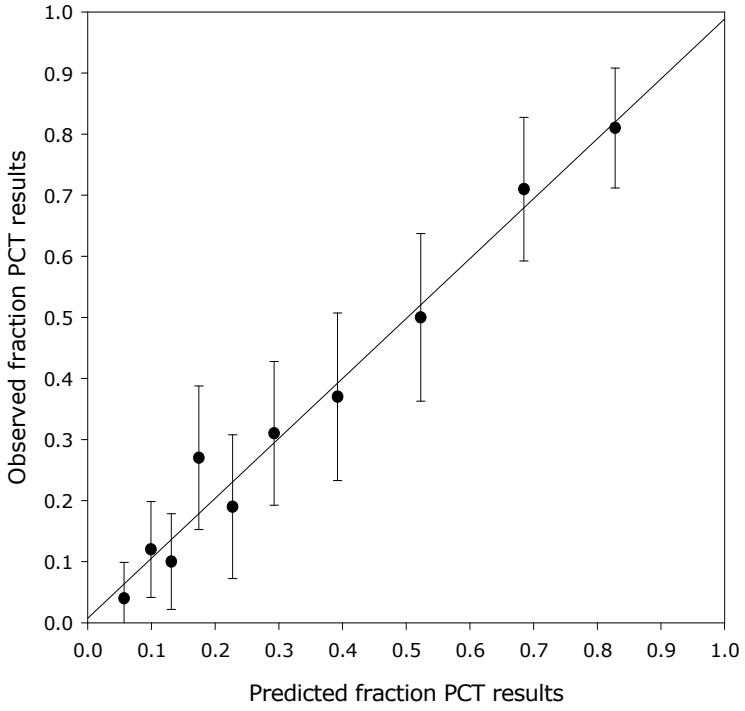


Figure 3 Calibration plot for the association between the chance of an abnormal PCT as predicted by the PCT model, and the mean observed test results.

Figure 3 shows the association between the probability of an abnormal PCT, as calculated from our PCT prediction model, and the observed test results. All points are located close to the line of equality ($x = y$), indicating a good calibration of the logistic model. This was confirmed by Hosmer-Lemeshow test statistic ($p = 0.73$).

Finally, we assessed the prognostic performances of the three pregnancy prediction models. Pregnancy model I, i.e. the model without the PCT, had an AUC of 0.61 (95% CI: 0.55-0.67). Pregnancy model II, i.e. the model with the PCT, had an AUC of 0.65 (95% CI: 0.59-0.71). This AUC was statistically significantly better (the difference in AUCs is 0.04; 95% CI: 0.004 – 0.07).

Pregnancy model III, including the PCT results only in those couples that had a chance of an abnormal PCT between 20% and 60%, had an AUC of 0.64 (95% CI: 0.58-0.70). The improvement between pregnancy model I and pregnancy model III was borderline statistically significant (difference in AUCs is 0.03; 95% CI: -0.006 – 0.06). The AUC of pregnancy model III was comparable with that of pregnancy model II, with a non-statistically significant difference (difference in AUCs is -0.01; 95% CI: -0.03 – 0.01).

Figure 4a shows the comparison between the probability of treatment-independent pregnancy predicted from pregnancy model I with pregnancy model III. Couples with an abnormal PCT had a poorer prognosis calculated with pregnancy model III than with pregnancy model I. The false-positive predicted PCTs are PCTs that are thought to be abnormal from the PCT model, but turn out to be normal when actually performed. Similarly, false-negative predicted PCTs are PCTs thought to be normal from the PCT model, but appear to be abnormal when performed in reality. The false-positive and false-negative predicted PCTs are spread equally over the scatter plot. Figure 4b shows the comparison between the chances on treatment-independent pregnancy predicted from pregnancy model II with pregnancy model III. For almost all couples the probabilities of a treatment-independent pregnancy at 6 months are equal in model II and model III, except for the few outliers. These outliers are due to the false-positive and false-negative predicted PCTs.

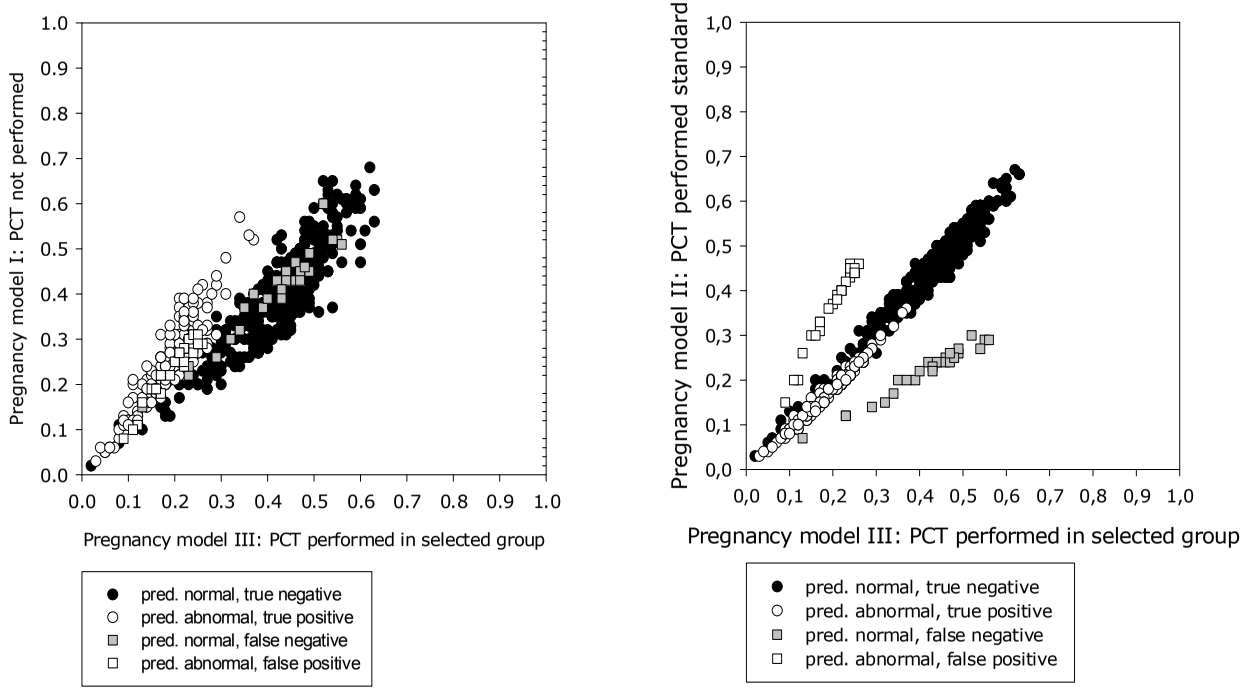


Figure 4 (**a**) Scatter plot in which the chance to conceive treatment-independent, in a period of 6 months, predicted with pregnancy model III is compared with pregnancy model I (pred. = predicted). (**b**) Scatter plot in which the chance to conceive treatment-independent, in a period of 6 months, predicted with pregnancy model III is compared with pregnancy model II.

Discussion

This study shows that medical history and semen analysis can predict the result of the PCT in approximately 50% of subfertile couples with a regular cycle. For the other 50% of the couples, i.e. patients with a chance of an abnormal PCT between 20% and 60%, the result of the PCT cannot be predicted reliably. Performance of the PCT might therefore have clinical relevance in these patients. We showed that replacing the result of the actual PCT by the predicted result of the PCT, in patients with a high or a low chance of an abnormal PCT, does not compromise the capacity to predict pregnancy from the basic fertility work-up. On the other hand, complete omission of the PCT from the prediction model results in a decreased prognostic performance.

Our study has some limitations. First, the data set of Snick was collected more than ten years ago. Since then, treatment has improved and the characteristics of the population have changed, with women starting to conceive at older age, and couples contacting fertility clinics at an earlier stage of subfertility. A second limitation is the fact that only one gynaecologist performed all PCTs in a rather homogeneous population. Consequently, in a clinic with a more heterogeneous population or in a clinic in which the PCT is performed in a less standardized setting, the capacity of the PCT prediction model might decrease. The problem of standardization of the PCT has been described before (Oei *et al.,* 1995b). In view of these limitations, we recommend that our PCT model should be evaluated in other populations.

A strength of this study is that application of the PCT model will lead to a reduction in the number of PCTs performed in clinical practice. With the prediction of the PCT model, the clinician can decide whether or not to perform the PCT. For example, a couple with secondary subfertility and semen volume of 4 mL, and a sperm concentration of 85×10^6/mL, 10% of which are progressively motile and with a normal morphology, will have a chance having an abnormal PCT of 19%. Therefore, the result of the PCT will most probably be normal and offering the PCT to such a couple is not likely to have additional value. Consequently, a primary subfertile couple with a semen volume of 4 mL, sperm concentration of 10×10^6/mL, of which 30% are progressively motile and 30% have normal morphology, will have a chance of an abnormal PCT of 51%. Therefore, the result of the PCT cannot be predicted reliably (probability between 0.2 and 0.6). The performance of the PCT is of additive value for this couple. Thus, in this way the clinician can decide, based on the PCT prediction, for which couples the performance of the PCT is of clinical relevance and for which it is not.

In this study, we have chosen cut-off values of 20% as the upper level for normal result of the PCT results and 60% as the lower border for abnormal PCT results. We have set the cut-off values at 20% and 60% supported by the thought that below a chance of 20% on an abnormal PCT, more than five PCTs should be performed to detect one abnormal PCT (in our data, 13% of the patients in this group indeed had an abnormal PCT). In the opposite way, with a chance of more than 60% on an abnormal PCT three out of five PCTs will be abnormal (in our data, 76% of the patients in this group indeed had an abnormal PCT). With these cut-off values, the false-negative fraction will be twice as small as the false-positive fraction. False-positive tests results can lead to overtreatment, while false-negative test results can lead to undertreatment. Therefore, these cut-off values for the PCT model result in a lower number of overtreated patients in comparison with the number undertreated patients. Only when the exact treatment for patients with an abnormal PCT is known, together with its side effects and costs, can a more considered choice for the cut-off values be made. These cut-off values might be helpful thresholds guiding future trial design.

Snick *et al*. reported a better performance of his prognostic models than we found in the present study; for the model with the PCT, we found an AUC of 0.65 (prediction of treatment-independent pregnancy at 12 months) whereas Snick *et al*. reported 0.79. For the model without the PCT, we found the AUC to be 0.61 versus 0.76 reported by Snick *et al*. This difference might be explained by the fact that Snick *et al.* used all 726 patients, whereas we excluded women with ovulation disorders. Another explanation might be that we considered the data as censored survival data, whereas Snick *et al.*, using a more conservative approach, considered pregnancy as a dichotomous event. We used a method recently introduced by Heagerty *et al.* (Heagerty *et al.,* 2000). The comparison of the method of Heagerty with the more conservative approach showed a decrease of 10% in the AUC of the three pregnancy models.

The effectiveness of the PCT has been assessed in a randomized clinical trial, which compared a strategy in which all couples had a PCT with a strategy in which none of the couples had a PCT (Oei *et al.,* 1998). At first, this type of study is considered to result in the standard of evidence. This trial attempted to replicate the imperfections of clinical practice in order to be relevant (Smith *et al.,* 2003). However, methodologically, this clinical management trial has been criticized by several authors (Cohlen *et al.,* 1999; Hull and Evers 1999; Hendry 1999). It has been stated that the included sample was not sufficiently exclusive (20% ovulatory disorder), that the intervention was used incompletely (36% missing PCTs in the

intervention group) and that a specific response to the test results was lacking. These methodological problems might have influenced the results of the trial (Bossuyt *et al.,* 2000). Furthermore, the occurrence of multiple pregnancies was not studied and treatment with IUI was always performed in a stimulated cycle. In contrast, our PCT model enables us to select a subgroup of patients in which the performance of the PCT might still be of use, thereby potentially identifying patients with a cervical factor in whom treatment with IUI without stimulation could prevent the occurrence of multiple pregnancies without compromising pregnancy rates.

Based on our findings, we suggest that a rearrangement of the basic fertility work-up is worthy of further evaluation. After taking patients' history and performing semen analysis, a prediction of PCT outcome could be calculated. Depending on this prediction, the clinician could decide to plan the PCT afterwards or leave it out, instead of routinely performing it in all patients. Before such an approach can be implemented in clinical practice, two other conditions have to be fulfilled. First, the present model has to be validated in other populations. Secondly, it has to be established whether the PCT has value in the selection of patients that will benefit from IUI without ovarian hyperstimulation, and patients that benefit from IUI with hyperstimulation. In contrast to male factor subfertility, in which there is evidence that IUI without hyperstimulation is as effective as IUI with hyperstimulation, evidence for the use or non-use of ovarian hyperstimulation in IUI for cervical factor subfertility is lacking. If IUI without hyperstimulation were to give results comparable with IUI with hyperstimulation, the PCT can prevent couples from additional side effects, e.g. ovarian hyperstimulation syndrome and twin pregnancies, and reduce costs.

In conclusion, we found that the result of the PCT can be predicted with data from history and semen analysis in about half of the couples with a regular cycle. If this is confirmed at external validation, and if the PCT remains an important prognostic and diagnostic test, our findings suggest that a trial to evaluate the PCT in the basic fertility work-up would involve performance of a PCT planned from the history and semen analysis.

Acknowledgements

The authors thank Herman Snick, who kindly agreed with the use of his data for the present study. The current project has been funded by ZonMW, grant 945-12-002.

References

Bossuyt PM, Lijmer JG, and Mol BW. (2000) Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 356,1844-1847.

Chaffkin LM, Nulsen JC, Luciano AA, and Metzger DA. (1991) A comparative analysis of the cycle fecundity rates associated with combined human menopausal gonadotropin (hMG) and intrauterine insemination (IUI) versus either hMG or IUI alone. *Fertil Steril* 55,252-257.

Check JH, Bollendorf A, Zaccardo M, Lurie D, and Vetter B. (1995) Intrauterine insemination for cervical and male factor without superovulation. *Arch Androl* 35,135-141.

Cohlen BJ, Te Velde ER, and Habbema JDF. (1999) Postcoital testing. Postcoital test should be performed as routine infertility test. *BMJ* 318,1008-1009.

Eimers JM, Te Velde ER, Gerritse R, Van Kooy RJ, Kremer J, and Habbema JDF. (1994) The validity of the postcoital test for estimating the probability of conceiving. *Am J Obstet Gynecol* 171,65-70.

Friedman A, Haas S, Kredentser J, Stewart E, and Schiff I. (1989) A controlled trial of intrauterine insemination for cervical factor and male factor: a preliminary report. *Int J Fertil* 34,199-203.

Glazener CM, Ford WC, and Hull MG. (2000) The prognostic power of the post-coital test for natural conception depends on duration of infertility. *Hum Reprod* 15,1953-1957.

Griffith CS and Grimes DA. (1990) The validity of the postcoital test. *Am J Obstet Gynecol* 162,615-620.

Harrell FE, Jr., Lee KL, and Pollock BG. (1988) Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 80,1198-1202.

Heagerty PJ, Lumley T, and Pepe MS. (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56,337-344.

Hendry WF. (1999) Postcoital testing. Male partner should be assessed. *BMJ* 318,1007-1008.

Hull MG and Evers JL. (1999) Postcoital testing. Criterion for positive test was not given. *BMJ* 318,1007-1009.

Kirby CA, Flaherty SP, Godfrey BM, Warnes GM, and Matthews CD. (1991) A prospective trial of intrauterine insemination of motile spermatozoa versus timed intercourse. *Fertil Steril* 56,102-107.

Oei SG. (1998) The postcoital test: a controversial investigation. *Eur J Obstet Gynecol Reprod Biol* 77,123-124.

Oei SG, Bloemenkamp KW, Helmerhorst FM, Naaktgeboren N, and Keirse MJ. (1996) Evaluation of the postcoital test for assessment of 'cervical factor' infertility. *Eur J Obstet Gynecol Reprod Biol* 64,217-220.

Oei SG, Helmerhorst FM, Bloemenkamp KW, Hollants FA, Meerpoel DE, and Keirse MJ. (1998) Effectiveness of the postcoital test: randomised controlled trial. *BMJ* 317,502-505.

Oei SG, Helmerhorst FM, and Keirse MJ. (1995a) When is the post-coital test normal? A critical appraisal. *Hum Reprod* 10,1711-1714.

Oei SG, Helmerhorst FM, and Keirse MJ. (2001) Routine postcoital testing is unnecessary. *Hum Reprod* 16,1051-1053.

Oei SG, Keirse MJ, Bloemenkamp KW, and Helmerhorst FM. (1995b) European postcoital tests: opinions and practice. *Br J Obstet Gynaecol* 102,621-624.

Smith S, Pfeifer SM, and Collins JA. (2003) Diagnosis and management of female infertility. *JAMA* 290,1767-1770.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Steyerberg EW, Eijkemans MJC, and Habbema JDF. (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 52,935-942.

Te Velde ER, Van Kooy RJ, and Waterreus JJ. (1989) Intrauterine insemination of washed husband's spermatozoa: a controlled study. *Fertil Steril* 51,182-185.

chapter (11)

GENERAL DISCUSSION

General Discussion

This thesis reports on the use of models for the prediction of spontaneous pregnancy in the management of subfertile couples. Our main finding was that gynaecologists differ widely in estimating fertility prognoses in subfertile couples, but that existing prediction models, which are based on female age, duration of subfertility, fertility history, sperm motility, and referral status are valid and that they can be used in clinical practice. In the next paragraphs, the most relevant findings of this thesis and the practical implications are discussed. In addition, suggestions for future research are provided.

Until recently, the main emphasis in reproductive medicine has been on finding causal diagnoses of subfertility and accompanying fertility treatment. Examples are ovulation induction in women diagnosed with anovulation, tubal surgery or in vitro fertilization (IVF) and embryo transfer in women with bilateral tubal disease, or IVF with or without intracytoplasmic sperm injection (ICSI), with or without surgical sperm retrieval in couples with reduced semen quality or even azoospermia.
In many couples such a causal diagnosis cannot be found. They have to be classified as having unexplained subfertility, mild male subfertility, cervical factor subfertility, mild endometriosis, or one-sided tubal pathology.

In these couples assisted reproductive techniques such as intrauterine insemination (IUI) or IVF can be considered. As these interventions are expensive and not without side effects, they should only be offered to a couple if their success rates clearly exceed the probability of a spontaneous pregnancy in that couple. (Wasson *et al.,* 1985; Te Velde and Cohlen 1999). This means that the fertility work-up also has to include a prognostic dimension (Habbema *et al.,* 2004). To be able to make adequate and reliable predictions in clinical practice, formal prediction models, in which the contribution of each factor is quantified, were developed in the recent past.

The first prediction model for spontaneous pregnancy was developed in 1993 (Bostofte *et al.,* 1993). This model was based on a Danish cohort and included three variables: duration of subfertility, the sperm penetration test, defined as the ability of sperm to move through fresh hen egg white, and the female subfertility factor, categorized as no disorder, having an

ovulation or cervical disorder, an anatomic disorder, or a combination of the disorders. Since then, eight other prediction models have been published (Bahamondes *et al.,* 1994; Eimers *et al.,* 1994; Wichmann *et al.,* 1994; Collins *et al.,* 1995; Snick *et al.,* 1997; Hunault *et al.,* 2004). They are presented in tables 1 and 2.

Table 1 Prognostic variables per model.

No.	Variables	Bostofte	Bahamondes	Wichmann	Eimers	Collins	Snick I	Snick II	Hunault I	Hunault II	Total N
1	Female age		+	+	+	+		+	+	+	7
2	Duration of subfertility	+	+	+	+	+	+	+	+	+	9
3	Secondary subfertility		+		+	+		+	+	+	6
4	Referral status								+	+	2
5	Family history of man				+						1
6	Urethritis of men in history			+							1
7	Semen										8
	• Sperm morphology		+	+							• 2
	• Sperm motility			+	+				+	+	• 4
	• Male defect (WHO)					+		+			• 2
8	Sperm penetration test	+									1
9	Cervical factor	+			+		+			+	4
10	Cycle length		+								1
11	Endometriosis					+					1
12	Anovulation	+					+	+			3
13	Tubal pathology	+				+	+	+			4
14	Pelvic surgery		+								1

Table 2 Prediction models for spontaneous pregnancy.

						Internal validation		External validation		
Model	**Y, country**	**N**	**Predictive factors**	**HR**	**95% CI**	**AUC**	**95% CI**	**AUC**	**95 % CI**	**Calibration**
Bostofte	1993	321	Duration of subfertility (per y)	0.85	0.76-0.95	-	-	-	-	-
	Denmark		Sperm penetration test (normal)	1.51	1.17-1.95					
			Ovulation or cervical factor	0.68	0.41-1.1					
			Anatomic disorder	0.45	0.29-0.71					
			Ovulation and anatomic disorder	0.30	0.12-0.76					
Bahamondes	1994	559	Female age (per y)	0.90	-	-	-	-	-	-
	Argentina		Duration of subfertility (per y)	0.85	-					
			Secondary infertility of the couple	2.45	-					
			Sperm morphology	1.09	-					
			Normal cycle	2.35	-					
			Pelvic surgery	0.38	-					
Wichmann	1994	907	Female age (per y)	0.97	0.95-1.0	-	-	-	-	-
	Finland		Duration of subfertility (per y)	0.84	0.79-0.90					
			Urethritis of men in history	0.57	0.33–0.97					
			Sperm motility < 20% and Quality of motility < 2	0.16	0.06–0.43					
			Sperm morphology < 70%	0.78	0.58–0.96					

Table 2 continued						Internal validation		External validation		
Model	**Y, country**	**N**	**Predictive factors**	**HR**	**95% CI**	**AUC**	**95% CI**	**AUC**	**95 % CI**	**Calibration**
Eimers	1994	914	Female age (per y)	0.97	0.93-1.01	-	-	0.60[a]	0.58-0.63	Poor[a]
	NL		Duration of subfertility (per y)	0.89	0.82-0.96					
			Female secondary infertility	1.74	1.3-2.4					
			Family history man	0.69	0.43-1.09					
			Semen motility (per %)	1.013	1.015-1.021					
			PCT – non progressive	2.1	1.2-3.7					
			PCT – progressive	4.3	2.5-7.4					
Collins	1995	2198	Female age (< 30 /> 30)	1.5	1.05-2.2	0.59	0.56-0.63	0.64[a]	0.61-0.66	Poor[a]
	Canada		Duration of subfertility (< 36 /> 36 months)	1.7	1.1-2.5					
			Secondary infertility of the couple	1.8	1.2-2.7					
			Male defect	0.47	0.3-0.8					
			Endometriosis	0.39	0.18-0.85					
			Tubal defect	0.50	0.4-0.6					
Snick I	1997	724	Duration of subfertility (< 24 months)	1.49	1.07-2.10	0.79	0.75-0.83	0.59[a]	0.57-0.62	Poor[a]
	NL		Abnormal PCT	0.26	0.17-0.40					
			Ovulation defect	0.35	0.21-0.58					
			Tubal defect	0.14	0.06-0.33					

[a] As reported by Van der Steeg *et al.* Human Reproduction 2007

[b] As reported by Hunault *et al.* Human Reproduction 2005

Table 2 continued						Internal validation		External validation		
Model	**Y, country**	**N**	**Predictive factors**	**HR**	**95% CI**	**AUC**	**95% CI**	**AUC**	**95 % CI**	**Calibration**
Snick II	1997	724	Female age (< 30 years)	1.35	1.00-1.82	0.76	0.72-0.80	0.59[a]	0.57-0.62	Poor[a]
	NL		Duration (< 24 months)	1.50	1.06-2.11					
			Secondary infertility of the couple	1.53	1.13-2.08					
			WHO semen defect	0.59	0.44-0.78					
			Ovulation defect	0.35	0.21-0.59					
			Tubal defect	0.13	0.05-0.30					
Hunault I	2004	2459	Female age < 31 y (per y)	0.97	-	0.62	-	0.63[a]		Good[a]
	NL		Female age > 31 y (per y)	0.92	-					
			Duration of subfertility (per y)	0.83	-					
			Female primary subfertility	0.56	-					
			Semen motility (per %)	1.008	-					
			Referral status (tertiary care)	0.78	-					
Hunault II	2004	2459	Female age < 31 y (per y)	0.97	-	0.64	-	0.59[a]		Moderate[a]
	NL		Female age > 31 y (per y)	0.94	-			0.59	0.46-0.73[b]	
			Duration of subfertility (per y)	0.88	-					
			Female primary subfertility	0.64	-					
			Semen motility (per %)	1.008	-					
			Referral status (tertiary care)	0.79	-					
			Abnormal PCT	0.39	-					

The most recently developed model is the synthesis model, designed in the Netherlands prior to the writing of this thesis (Hunault *et al.,* 2004). This model was based on the patient data of the three previous cohorts of Eimers, Collins and Snick, combined into one. It includes the variables female age, duration of subfertility, subfertility being primary or secondary, referral status (being referred by the general practitioner or another gynaecologist), percentage motile sperm, and result of the postcoital test.

The performance of these prediction models was still unknown at the time the studies described in this thesis were developed, as they had not been validated in external populations. This is an extremely important issue, because most prediction models tend to be overoptimistic when applied to other populations than the one in which they were developed (Stolwijk et *al.*, 1996).

In this thesis, we report on the external validation of the recently developed synthesis prediction model for spontaneous pregnancy (model of Hunault) in a large prospective validation study. We found that this prediction model accurately predicted the chances of a spontaneous pregnancy among subfertile ovulatory couples.

We also showed that prediction models can be used to classify subfertile couples accordingly, and that such an approach can prevent overtreatment. In a randomized clinical trial that was performed recently and closely related to the research presented in this thesis, we found that, in couples suffering from unexplained subfertility and a 12-month prognosis for spontaneous pregnancy between 30% and 40%, a large benefit of treatment with IUI could be excluded (Steures *et al.,* 2005). We therefore advise to use prediction models as the final step of the fertility work-up, in order to counsel patients properly.
It may be that it is possible to improve the existing model by redefinition of the current prognosticators fertility history, semen analysis and postcoital test (chapters 6, 9, 10).

For example, female secondary subfertility is in the existing prediction model defined as having conceived before, regardless of the outcome of that pregnancy and whether it occurred in the current or in a previous partnership. In chapter 6, we showed that only previous intrauterine pregnancies in the current partnership affect the chance to conceive. The impact of previous pregnancies can thus be fully explained due to previous

pregnancies in current partnership, and not to pregnancies in previous partnerships.

Another remarkable finding was that couples in whom the men had fathered a child in a previous partnership were less likely to conceive in comparison with primary subfertile men. We hypothesize that this is due to negative selection bias. Men who fathered a child in a previous partnership have normal reproductive capacity. Thus, the fact that conception does not occur in their new relationship is probably due to female factors, some of which cannot be defined and therefore not controlled. The fact that a couple in which the woman had conceived in a previous partnership does not have a lower fertility chance indicates that the impact of the woman on fecundity is probably much larger than that of the man.

In chapters 7 and 8, we report on the studies to evaluate whether the addition of prognosticators BMI and FSH, which are not included in the synthesis model, can improve the existing model. Both variables appeared to be conditional predictors for spontaneous pregnancy and were subsequently entered in the model.

When we focused on FSH and evaluated in what proportion the addition of basal FSH to the Hunault prediction model altered the decision for treatment or expectant management, we observed that it did only leads to a change in treatment policy in less than 5% of all couples. Therefore, we do not recommend routine testing of basal FSH in subfertile couples. Whether the test is useful in a more selected group of subfertile women should be subject of further research, in which not only the effects but also costs should be considered.

The data generated in this thesis provide the opportunity to integrate the potential new factors, presented in chapters 6, 7, 8, 9 and 10 into a new prediction model. In a first analysis, we could improve the synthesis model for spontaneous pregnancy with the new factors cycle length, BMI, a history of ongoing pregnancy and history of spontaneous abortion, both in the current relationship (table 3). The semen parameters volume, sperm concentration, and WHO sperm morphology have additional prognostic value, next to sperm motility, which is the only sperm variable in the existing model. Basal FSH, fertility problems in the male's family, a history

of induced abortion or fertility treatment in the current relationship and a pregnancy in a previous relationship of the man were not selected.

The performance of the new model and the Hunault model were comparable, with respect to calibration and discrimination (AUC 0.61). However, in 17.5% of all couples, the treatment policy altered from treatment to expectant management and vice versa, when we used the new model. Whether such a shift in treatment policy will lead to more pregnancies for less costs has to be evaluated in clinical practice.

Table 3 Results of the univariable and multivariable Cox' proportional hazard analysis.

	Univariable analysis		**Multivariable analysis**	
	HR	**95% CI**	**HR**	**95% CI**
Female age (per year older)			0.97	0.95–0.99
Age < 31 (per year older)	0.98	0.94–1.02		
Age ≥ 31 (per year older)	0.97	0.93–1.01		
Duration of subfertility (per year longer)	0.80	0.72–0.88	0.84	0.76–0.93
Referral status (3rd care versus 2nd care)	0.45	0.28–0.73	0.46	0.29–0.75
Cycle length (per day longer)	1.08	1.04–1.13	1.08	1.04–1.12
Basal FSH day 2-4				
≤ 8.0 IU/L				
> 8.0 IU/L (per IU/L more)	0.97	0.92–1.03	-	
Body Mass Index (BMI)				
BMI < 21 kg/m^2 (per unit)	0.89	0.78–1.01	0.89	0.78–1.01
21 ≤ BMI < 29 kg/m^2	1		1	
BMI ≥ 29 kg/m^2 (per unit)	0.94	0.90–0.99	0.94	0.89–0.98
Previous live birth in the current partnership	1.60	1.32–1.95	1.67	1.37–2.05
Previous miscarriage in the current partnership	1.50	1.19–1.89	1.30	1.02–1.65
Previous induced abortion in the current partnership	1.10	0.57–2.12	-	
Previous pregnancy man in other partnership	0.77	0.54–1.11	-	
Previous pregnancy after ART in the current partnership	0.76	0.40–1.44	-	
Fertility problems in the male's family	0.94	0.79–1.11	-	
Semen volume				
≤ 2.0 mL (per mL less)	0.76	0.57–1.03	0.81	0.60–1.09
> 2.0 mL	1		1	
Semen concentration				
≤ 40x 10^6/mL (per 10x10^6 less)	0.91	0.85–0.98	0.94	0.87–1.02
> 40x 10^6/mL	1		1	
Semen motility (per 10% less)	0.94	0.90–0.98	0.96	0.92–0.99
Semen morphology				
≤ 20% normal (per 10% less)	0.87	0.73–1.03	0.87	0.73–1.05
> 20% normal	1		1	

Consequences for clinical practice

Based on the findings in this thesis, we conclude that the synthesis prediction model for spontaneous pregnancy makes valid predictions and should be implemented in clinical practice to counsel patients properly, thereby possibly preventing overtreatment.

At present, none of the guidelines on fertility care advise the use of prediction models in clinical decision-making (Bloomington (MN): Institute for Clinical Systems Improvement (ICSI) 2004; Boston (MA) 2003; National institute for clinical excellence 2004). Potential factors that may have hampered implementation so far are a general lack of knowledge about prediction models and doubts about the need for prediction models in reproductive medicine. In addition, the current organization of fertility care, the underlying incentive structures and healthcare insurer policies may play a role. Finally, couples can have a great sense of urgency and demand fertility treatment.

Implementation of clinical prediction rules could lead to a reduction of costs of treatment, reduction of practice variation and to a reduction of complications of treatment, of which multiple pregnancies is the most important one.

Of paramount importance is therefore the development of an implementation plan. But first, factors that facilitate or hamper the implementation of tailored expectant management by prediction models in fertility care should be identified. This knowledge can then be used to develop an effective implementation strategy.

In conclusion, we have shown that prediction models for spontaneous pregnancy are needed in clinical practice. The models are fairly robust and will help both the subfertile couple and the doctor to decide to start fertility treatment or not. Refinement of these prognostic models and addition of new potential prognosticators may improve existing prediction models. Implementation programs are warranted to actually make prediction models standard practice in reproductive medicine.

References

Bahamondes L, Alma FA, Faundes A, and Vera S. (1994) Score prognosis for the infertile couple based on historical factors and sperm analysis. *Int J Gynaecol Obstet* 46,311-315.

Bloomington (MN): Institute for Clinical Systems Improvement (ICSI). (2004) Diagnosis and management of basic infertility. www.guideline.gov.

Bostofte E, Bagger P, Michael A, and Stakemann G. (1993) Fertility prognosis for infertile couples. *Fertil Steril* 59,102-107.

Boston (MA). (2003) Infertility. A guide to evaluation, treatment, and counseling. www.guideline.gov.

Collins JA, Burrows EA, and Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril* 64,22-28.

Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CW, and Habbema JDF. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 61,44-52.

Habbema JDF, Collins J, Leridon H, Evers JLH, Lunenfeld B, and Velde ERT. (2004) Towards less confusing terminology in reproductive medicine: a proposal. *Fertility and Sterility* 82,36-40.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JL, and Te Velde ER. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 19,2019-2026.

National institute for clinical excellence. (2004) Fertility guideline: assessment and treatment for people with fertility problems. www.nice.org.uk/pdf/download.aspx?o=CG011fullguideline.

Snick HK, Snick TS, Evers JL, and Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod* 12,1582-1588.

Steures P, Van der Steeg JW, Hompes PGA, Bossuyt PMM, Van der Veen F, Habbema JDF, Eijkemans MJC, Broekmans FJ, Verhoeve HR, and Mol BWJ. A randomised controlled trial comparing intrauterine insemination with controlled ovarian hyperstimulation versus expectant management in couples with unexplained subfertility and an intermediate prognosis. Human Reproduction 20, i113. 2005.

Stolwijk AM, Zielhuis GA, Hamilton CJ, Straatman H, Hollanders JM, Goverde HJ, Van Dop PA, and Verbeek AL. (1996) Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. *Hum Reprod* 11,2298-2303.

Te Velde ER and Cohlen BJ. (1999) The management of infertility. *N Engl J Med* 340,224-226.

Wasson JH, Sox HC, Neff RK, and Goldman L. (1985) Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 313,793-799.

Wichmann L, Isola J, and Tuohimaa P. (1994) Prognostic variables in predicting pregnancy. A prospective follow up study of 907 couples with an infertility problem. *Hum Reprod* 9,1102-1108.

chapter (12)

SUMMARY, CONCLUSIONS & RECOMMENDATIONS

Samenvatting, conclusies & aanbevelingen

Summary

This thesis reports on the ability to prediction spontaneous pregnancy chances in couples that have been trying to conceive unsuccessfully for at least one year. The clinical value and prognostic capacity of several factors, that are investigated during the routine fertility work-up, were evaluated both in a theoretical and practical setting. The Collaborative Effort for Clinical Evaluation in Reproductive Medicine group (CECERM) collected 7860 consecutive subfertile couples prospectively in 38 hospitals in the Netherlands between January 1, 2002 and February 1st, 2004.

In the introduction the fertility work-up of subfertile couples is outlined, and the most relevant prediction models for spontaneous pregnancy are resumed. Chapter 2 focuses on at what moment to perform the basic fertility work-up. In chapters 3 and 4 the ability of gynaecologist to predict pregnancy chances is evaluated. In chapter 5 the prediction model for spontaneous pregnancy of Hunault is validated. Chapters 6 to 10 report on the clinical impact and predictive capacity of existing and new factors of the fertility work up, i.e. previous pregnancies, semen, PCT, obesity and basal FSH.

In **chapter 1** we describe the fertility work-up of subfertile couples. After completion of the fertility work-up, couples can be distinguished in couples with a causal diagnosis, in whom treatment is mandatory, and couples in whom no diagnosis could be made. In the later group the prognosis will determine whether treatment is indicated. We hypothesize that prediction models might be an essential tool to calculate this prognosis. The most relevant prediction models for spontaneous pregnancy are summarized in short.

Chapter 2 demonstrates why women younger than 39 years and with a regular cycle, should not have a basic fertility work-up earlier than after 12 months of child wish. Four reasons are postulated. Couples, that tried to conceive for only six months without success, still have a chance to conceive without treatment of 50%. Secondly, these couples have a lower prevalence of fertility disorders compared with subfertile couples, leading to an increase in false positive diagnoses. Thirdly, fertility treatment will have less additional effects in these couples, and fourthly, postponement of treatment

in such couples will prevent complications as ovarian hyperstimulation syndrome and multiple pregnancies.

Chapter 3 reports on a survey among Dutch gynaecologists who estimated pregnancy probabilities (without treatment, with IUI, with IVF) and proposed therapeutic regimens for 16 fictive couples, i.e. expectant management, IUI or IVF. The gynaecologists differed widely in estimating pregnancy chances (interclass correlation coefficient [ICC]: 0.34). Furthermore there was a huge variation in the proposed therapeutic regimens (Kappa [κ]: 0.21). When prediction models were used, the concordance (κ) for treatment decisions increased from 0.21 to 0.38. The number of gynaecologists counselling for expectant management increased from 39% to 51%, whereas counselling for IVF dropped from 23% to 14%.

Chapter 4, based on the same survey, determines the importance of each individual factor of the fertility work-up in the subsequent management of subfertile patients. Female age, duration of subfertility and FSH-level were the main factors in clinical decision-making of gynaecologists in subfertility. Gynaecologists overestimated the importance of basal FSH-level and underestimated the importance of previous pregnancy in clinical decision-making as compared with their importance as reported in prediction models and guidelines.

Chapter 5 describes performance of the prediction model for spontaneous pregnancy of Hunault in a general subfertile population in terms of accuracy and discriminative capacity. The prediction model was evaluated among 3021 subfertile couples enrolled in 38 hospitals, of whom 537 (18%) had a spontaneous ongoing pregnancy, 55 (2%) a non-successful pregnancy, 1316 (44%) started treatment, 820 (27%) neither started treatment nor became pregnant, and 280 (9%) were lost to follow-up. Calibration of the prediction model was almost perfect. For couples with a prognosis ≥ 40%, the cumulative ongoing pregnancy rate without treatment at 12 months was 46%. The discriminative capacity was similar to the one in which the model was developed (c-statistic: 0.59).

Chapter 6 reports on the capacity of a detailed pregnancy history to predict spontaneous ongoing pregnancy in subfertile couples. Of 4445 included couples, that completed a fertility work-up, 793 (18%) had a spontaneous ongoing pregnancy within one year of follow-up. Previous live birth and miscarriage in current partnership were both associated with higher

fecundity as compared with primary infertility (Hazard rate ratios [HR] 1.4 and 1.3, respectively). Pregnancies in women's previous partnerships did not affect the fecundity of the couple. A pregnancy in previous partnership of the male partner was associated with lower fecundity (HR 0.76). A previous pregnancy after fertility treatment was also associated with lower fecundity (HR 0.52).

Chapter 7 reports on how obesity affects the chance of a spontaneous pregnancy in ovulatory women presenting with subfertility. We included 3029 consecutive subfertile couples without anovulation, double-sided tubal pathology, or severe semen abnormalities. Fertility rates declined linearly with a body mass index (BMI) over 29 kg/m^2. Women with a high BMI had a 5% lower fecundity per kg/m^2 increase (HR 0.95), even after adjustment for possible related factors, i.e. female age, duration of subfertility, previous pregnancies, semen analysis, current smoking and being referred by a general practitioner or specialist.

Chapter 8 reports the clinical impact and predictive capacity of basal FSH on spontaneous ongoing pregnancy in subfertile ovulatory women. We included 3519 ovulatory subfertile women, without two-sided tubal pathology, or severe semen abnormalities. Basal FSH levels of 8 IU/L or higher were associated with a decreased probability of spontaneous ongoing pregnancy (HR 0.93 per IU/L). In a multivariable analysis, female age (HR 0.97 per year), cycle length (HR 0.96 per day) and FSH levels ≥ 8 IU/L (HR 0.93 per IU/L) were negative predictors for spontaneous ongoing pregnancy. The number of couples in whom the FSH level altered management decisions was low. Routine testing of basal FSH in subfertile couples was not recommended.

In **chapter 9** the WHO criteria validated in a prospective multicenter cohort study in 3345 consecutive subfertile couples of whom 517 had a spontaneous ongoing pregnancy. According to the WHO criteria for semen, only 11% of the men had normozoospermia. The cumulative pregnancy rate in men with normozoospermia did not differ significantly from the rate in men with a semen defect (28% versus 23%; $p = 0.21$). We observed lower probabilities of fathering a child up for sperm concentration up to 40×10^6/mL, for sperm motility over the whole range, and for sperm morphology up to 20% normophorms, in contrast to the WHO cut-off values of 20×10^6/mL, 50% and 15%, respectively. A prediction model based on the redefined semen parameters was able to make a finer differentiation

between subfertile men with a low and a higher probability of fathering a child (16 to 31%). It was concluded that the current WHO criteria for semen quality were not able to identify couples who are likely to conceive without treatment in a subfertile population.

Chapter 10 investigates whether medical history and semen analysis could predict the result of the PCT and evaluated the capacity of the PCT to predict spontaneous ongoing pregnancy. We evaluated 522 couples, previously reported on by Snick *et al.*, of whom 179 (34%) had an abnormal PCT. The PCT could be predicted accurately with a model including previous pregnancy (odds ratio [OR] 2.1), semen volume (OR 0.88), sperm concentration (OR 0.96), sperm motility (OR 0.97) and sperm morphology (OR 2.7). The area under the ROC-curve of the model was 0.81. In order to predict spontaneous ongoing pregnancy, the result of the actual PCT could be replaced by the predicted result of the PCT in about half of the couples, without compromising its predictive capacity.

Conclusions and recommendations

- Gynaecologists are not able to estimate probabilities of pregnancy in a reliable and reproducible way in subfertile couples. Their therapeutic regimens for fertility patients show huge variation.

- The prediction model of Hunault calculates the chance of a spontaneous ongoing pregnancy among subfertile couples accurately. We advise to use this or a comparable model to predict spontaneous pregnancy as a final and essential step of the basic fertility work-up in order to decide for fertility treatment or expectant management.

- The capacity of pregnancy history to predict a spontaneous pregnancy depends on the mode of conception (spontaneous or with ART), the localisation of the pregnancy (intrauterine or ectopic) and whether the pregnancy occurred in current partnership or not.

- The result of the PCT can be predicted in approximately 50% of the subfertile couples, without compromising its capacity to predict pregnancy. We suggest performing the PCT only in couples with primary subfertility in whom the man has good semen quality.

- The current WHO criteria for semen quality are of limited use in a large subfertile population in identifying couples who were likely to conceive without medical assistance. They should be replaced by new criteria for semen quality.

- Basal FSH contributes to the prediction of pregnancy. Its role is limited for treatment decisions in a general subfertile population. We advise to reserve the basal FSH test for women between 35 and 38 years, women with a shortening of the cycle length and women with a family history with premature ovarian failure.

- Obesity is an important risk indicator for a lower chance of conception in subfertile women with an ovulatory cycle.

Samenvatting

Subfertiliteit wordt gedefinieerd als ongewenste kinderloosheid na ten minste 1 jaar onbeschermde coïtus. Dit overkomt ongeveer een op de tien paren met actieve kinderwens. Jaarlijks worden hierdoor in Nederland bijna 20 000 paren geconfronteerd met subfertiliteit.

De meerderheid van deze subfertiele paren zal medische hulp zoeken bij fertiliteitscentra. Daar zal met een oriënterend fertiliteitsonderzoek (OFO) worden gestart om de oorzaak van de subfertiliteit te achterhalen. Het huidige OFO bestaat uit verschillende testen. Het begint met een gedetailleerde anamnese, gevolgd door het testen van de ovulatie, het zaadonderzoek, de samenlevingstest, waarin 12 uur na de samenleving wordt gekeken of er nog levend zaad aanwezig is in het slijm van de vrouw, en tot slot wordt de doorgankelijkheid van de eileiders getest. Nadat het OFO is afgerond, kan een diagnose worden gevonden, zoals anovulatie, afgesloten eileiders, endometriose, de cervix factor, mannelijke factor of een stoornis in de seksuele functie. Echter, in 20% van de paren wordt geen oorzaak voor de ongewenste kinderloosheid gevonden.

Bij paren, waarbij de kans op zwangerschap ernstig verstoord wordt door de gevonden diagnose, is het logisch dat behandeling noodzakelijk is. Bijvoorbeeld bij een paar dat subfertiel is, omdat de vrouw anovulatoir is, zal worden begonnen met een behandeling met ovulatie-inductie, waarbij de eisprong wordt aangestuurd. Bij paren waarvan de vrouw is gediagnosticeerd met dubbelzijdig afgesloten eileiders zal begonnen worden met tubachirurgie of met invitro fertilisatie (IVF), afhankelijk van de schade aan de eileiders. Bij paren waarbij gevonden is dat de man een zeer slechte zaadkwaliteit heeft, zal de behandeling bestaan uit IVF met of zonder intracytoplasmatische sperma injectie (ICSI).

Bij paren waarbij geen oorzakelijke diagnose is gevonden, is de resterende kans op een spontane zwangerschap van groot belang om te bepalen of fertiliteitsbehandeling al dan niet nodig is. Dit is bijvoorbeeld het geval bij paren met onverklaarde subfertiliteit, bij paren waarbij de man een matig afwijkende zaadkwaliteit heeft, en bij paren waarbij de vrouw is gediagnosticeerd met een cervix factor, milde endometriose of eenzijdige afwijking aan de eileiders.

Fertiliteitsbehandelingen moeten namelijk alleen dan worden verricht als de kans op zwangerschap hierdoor duidelijk groter is dan zonder behandeling. Daarom is het uitermate belangrijk paren te onderscheiden die een geringe kans hebben op een spontane zwangerschap, bij wie intra-

uteriene inseminatie (IUI) of IVF noodzakelijk is, van paren die nog steeds een goede kans op een spontane zwangerschap hebben. Prognostische modellen voor zwangerschap kunnen van grote waarde zijn om deze kansen in te schatten. Op deze manier kunnen mogelijk onnodige fertiliteitsbehandelingen worden voorkomen.

In dit proefschrift wordt beschreven in hoeverre het op dit moment mogelijk is om bij paren met onvervulde kinderwens de kans te voorspellen om zwanger te worden zonder fertiliteitsbehandeling. De klinische relevantie en de voorspellende waarde van een aantal factoren, die onderdeel zijn van het oriënterend fertiliteitsonderzoek (OFO), werden geëvalueerd in een theoretische en klinische setting. In de periode van januari 2002 tot en met 1 februari 2004 werden in 38 Nederlandse ziekenhuizen 7 860 subfertiele paren verzameld in het Oriënterend Fertiliteits Onderzoek project (OFO-project, internationaal ook bekend als the Collaborative Effort for Clinical Evaluation in Reproductive Medicine group (CECERM)).

In de inleiding van dit proefschrift beschrijven wij de opbouw van het standaard oriënterend fertiliteitsonderzoek. Verder bespreken wij de meest relevante predictiemodellen voor het voorspellen van spontane zwangerschap. Hoofdstuk 2 richt zich op de vraag wat het beste moment is om het OFO in gang te zetten. In de hoofdstukken 3 en 4 evalueren wij in hoeverre gynaecologen in staat zijn om zelf bij subfertiele paren de kans op een zwangerschap te voorspellen. In hoofdstuk 5 valideren we het predictiemodel voor spontane zwangerschap van Hunault. In de hoofdstukken 6 tot en met 10 worden de klinische relevantie en prognostische waarde beschreven van bestaande en nieuwe onderdelen van het OFO; de obstetrische voorgeschiedenis, het zaadonderzoek, de samenlevingstest, ook wel de postcoitumtest (PCT) genoemd, obesitas en het basaal follikel stimulerend hormoon (FSH).

Hoofdstuk 1 beschrijft het oriënterend fertiliteitsonderzoek. Nadat het OFO is afgerond kunnen subfertiele paren worden onderscheiden in paren met een oorzakelijke diagnose, bij wie behandeling duidelijk nodig is, en paren bij wie geen reden voor de onvruchtbaarheid is gevonden. In deze laatste groep is de kans op een spontane zwangerschap van doorslaggevende betekenis, waarmee kan worden bepaald of fertiliteitsbehandelingen al geïndiceerd zijn of niet. Onze hypothese is dat predictiemodellen een belangrijke rol kunnen spelen om deze kans te berekenen. De meest

relevante predictiemodellen voor spontane zwangerschap worden in het kort samengevat.

Hoofdstuk 2 toont aan waarom het OFO bij vrouwen onder de 39 met een regelmatige cyclus pas moet worden gestart na 12 maanden onvervulde kinderwens en niet eerder. Dit wordt onderbouwd met vier argumenten. Ten eerste hebben paren, die slechts zes maanden onvervulde kinderwens hebben, altijd nog een kans van meer dan 50% om zwanger te worden zonder fertiliteitsbehandeling. Ten tweede is de prevalentie van vruchtbaarheidsafwijkingen lager bij deze groep mensen dan bij paren, die al meer dan een jaar onvervulde kinderwens hebben. Dit zal resulteren in een toename van fout positieve diagnoses. Ten derde zal bij deze paren het toegevoegde effect van fertiliteitsbehandeling gering zijn. Tot slot voorkomt bij deze paren het uitstellen van behandeling complicaties zoals het ovariële hyperstimulatie syndroom (OHSS) en het krijgen van een meerling-zwangerschap.

Hoofdstuk 3 beschrijft de resultaten van een enquête onder Nederlandse gynaecologen, waarin zij gevraagd werden om in 16 fictieve casus de kans op zwangerschap in te schatten (zonder behandeling, met intra-uteriene inseminatie (IUI) en met IVF) en vervolgens om voor elk van deze 16 paren een beleid voor te stellen, te weten: afwachtend beleid, IUI of IVF. De gynaecologen verschilden onderling aanzienlijk in hun voorspellingen (interclass correlation coëfficiënt [ICC]: 0,34). Verder bestond er een enorm verschil tussen hun onderlinge beleidsvoorstellen (Kappa [κ]: 0,21). Op het moment dat zij gebruik konden maken van predictiemodellen werden gynaecologen het vaker eens over het te volgen beleid (de concordantie (κ) nam toe van 0,21 naar 0,38). Het aantal gynaecologen dat afwachtend beleid voorstelde nam toe van 39% naar 51%, terwijl het aantal dat IVF voorstelde, afnam van 23% naar 14%.

In **hoofdstuk 4**, dat gebaseerd is op de enquête uit hoofdstuk 3, analyseren wij het individuele belang van elke factor van het oriënterend fertiliteitsonderzoek in relatie tot het vervolgbeleid van subfertiele paren. Gynaecologen baseerden hun klinische beleid vooral op de leeftijd van de vrouw, de duur van onvervulde kinderwens en de hoogte van het FSH. Gynaecologen overschatten het belang van FSH en onderschatten de waarde van de obstetrische voorgeschiedenis in hun voorstellen voor fertiliteitsbehandelingen in vergelijking met wat er in de literatuur

gerapporteerd wordt over de waarde van FSH en de obstetrische voorgeschiedenis.

In **hoofdstuk 5** wordt beschreven hoe goed het predictiemodel van Hunault de kans op een spontane zwangerschap voorspelt in een algemene subfertiele populatie in termen van accuratesse en discriminerend vermogen. Hiervoor werd het predictiemodel geëvalueerd op 3 021 subfertiele paren die werden verzameld in 38 Nederlandse ziekenhuizen. Van deze paren kregen 537 (18%) een spontane doorgaande zwangerschap, 55 (2%) een miskraam, 1 316 paren (44%) begonnen met een fertiliteitsbehandeling, 820 (27%) waren in een jaar nog niet zwanger, maar ook nog niet begonnen met een behandeling. Ten slotte was van 280 paren (9%) de afloop aan het eind van de studie niet bekend. Calibratie van het predictie model was bijna perfect; in de groep patiënten met een prognose op een spontane zwangerschap van ≥ 40% was het cumulatieve doorgaande zwangerschap percentage na een jaar 46%. Het discriminerende vermogen van het predictie model was vergelijkbaar in deze populatie met de populatie waarmee het model op was ontwikkeld (c-statistic: 0,59).

In **hoofdstuk 6** wordt gerapporteerd over de waarde van een gedetailleerde obstetrische voorgeschiedenis in relatie tot het voorspellen van de kans op een spontane doorgaande zwangerschap bij subfertiele paren. Van de 4 445 geïncludeerde paren die het OFO hadden voltooid, werden 793 (18%) spontaan zwanger van een doorgaande zwangerschap in een jaar. Paren met een levend geborene of miskraam in de huidige relatie hadden een grotere kans op een spontane zwangerschap in vergelijking met paren die nog niet eerder zwanger waren (Hazard rate ratios [HR] 1,4 en 1,3, respectievelijk). Eerdere zwangerschappen van de vrouw in een eerdere relatie waren niet van belang voor de vruchtbaarheid. Als de man een zwangerschap had verwekt in een eerdere relatie, dan was dat geassocieerd met een lagere kans op zwangerschap in de huidige relatie (HR 0,76). Een eerdere zwangerschap ontstaan na fertiliteitsbehandeling was ook geassocieerd met een lagere kans op zwangerschap in de huidige relatie (HR 0,52).

In **hoofdstuk 7** wordt beschreven in hoeverre obesitas de kans op een spontane zwangerschap beïnvloedt bij subfertiele vrouwen met een ovulatoire cyclus. In totaal werden 3 029 subfertiele paren zonder de diagnose anovulatie, tweezijdige tubapathologie of ernstige mannelijke

factor geïncludeerd. De kans op een spontane zwangerschap daalde recht evenredig met de body mass index (BMI) vanaf 29 kg/m^2. Obese vrouwen (BMI > 29) hadden per eenheid toename van de BMI een 5% lagere kans op zwangerschap (HR 0,95), zelfs na correctie voor mogelijk beïnvloedende factoren zoals leeftijd van de vrouw, duur van kinderwens, eerdere zwangerschappen, het zaadonderzoek, rookgewoontes en de verwijsstatus (verwezen door huisarts of specialist).

In **hoofdstuk 8** vindt u het onderzoek naar de klinische relevantie en prognostische waarde van het basaal FSH in relatie tot de kans op een spontane zwangerschap bij subfertiele vrouwen met een ovulatoire cyclus. In totaal werden 3 519 subfertiele vrouwen met een ovulatoire cyclus, zonder tweezijdige tubapathologie en zonder dat de man was ernstige mannelijke factor had, geïncludeerd. Waarden van het basaal FSH boven de 8 IU/L waren geassocieerd met een lagere kans op een spontane doorgaande zwangerschapwas (HR 0,93 per IU/L). In een multivariabele analyse werd ook gevonden dat leeftijd van de vrouw (HR 0,97 per jaar ouder), cyclusduur (HR 0,96 per dag korter) naast een FSH-waarde ≥ 8 IU/L (HR 0,93 per IU/L) negatieve voorspellers waren voor de kans op een spontane doorgaande zwangerschap. Het aantal paren waarin de waarde van het FSH daadwerkelijk van invloed zou zijn op het te volgen fertiliteitsbeleid was gering. Routinematig testen van het basaal FSH bij alle subfertiele paren werd om die redenen niet geadviseerd.

In **hoofdstuk 9** worden de World Health Organisation (WHO) criteria voor zaadanalyse gevalideerd in een prospectief multicenteronderzoek onder 3 345 consecutieve subfertiele paren. Van hen werden 517 paren binnen een jaar zwanger van een spontane doorgaande zwangerschap. Volgens de WHO-criteria voor zaadanalyse viel slechts 11% van de mannen binnen de groep normozoöspermie. Het cumulatieve zwangerschapspercentage in de groep mannen met normozoöspermie verschilde niet significant van mannen met een afwijkende zaadanalyse (28% tegen 23%; p = 0,21). Wij vonden een lagere kans op het verwekken van een zwangerschap bij paren waarvan de man een zaadconcentratie onder de 40 x 10^6/ml had, of waarvan de morfologie van het zaad onder de 20% normaal was. Dit in tegenstelling tot de WHO, die afkapwaarden hanteert van 20 x 10^6/ml voor zaadconcentratie en 15% voor het percentage zaad met een normale vorm. Met een predictie model dat gebaseerd was op de geherdefinieerde zaadparameters kon een preciezer onderscheid worden gemaakt tussen mannen met een lage en een hogere kans op het verwekken van een zwangerschap (van 16% tot 31%).

Concluderend: de huidige WHO-criteria voor zaadanalyse zijn in een algemene subfertiele populatie niet bruikbaar om paren te identificeren die een grote kans hebben op een spontane zwangerschap, waarbij geen fertiliteitsbehandeling nodig is.

In **hoofdstuk 10** wordt onderzocht of met de anamnese en het zaadonderzoek de uitkomst van de postcoitumtest (PCT) kan worden voorspeld. Verder wordt de prognostische waarde van de PCT om de kans op spontane zwangerschap te voorspellen geëvalueerd. Van de 522 paren waar Snick *et al.* eerder over schreven, hadden 179 paren (34%) een afwijkende PCT. Het resultaat van de PCT kon accuraat worden voorspeld met een model bestaande uit de factoren obstetrische voorgeschiedenis (odds ratio [OR] 2,1), zaadvolume (OR 0,88), zaadconcentratie (OR 0,96), motiliteit van het zaad (OR 0,97) en de morfologie van het zaad (OR 2,7). Het model had een 'area under the ROC-curve' van 0,81. Zelfs als in 50% van de paren de uitkomst van de PCT berekend zou worden in plaats van daadwerkelijk getest, zou dat niet ten koste gaan van het prognostische vermogen van de PCT om de kans op een spontane zwangerschap te berekenen.

Conclusies en aanbevelingen

- Gynaecologen zijn niet in staat om zwangerschapskansen voor subfertiele paren op een betrouwbare en reproduceerbare manier in te schatten. Zij verschillen onderling aanzienlijk van mening ten aanzien van de door hen voorgestelde fertiliteitsbehandelingen.

- Met het predictiemodel van Hunault kan voor subfertiele paren de kans op een spontane zwangerschap nauwkeurig worden berekend. Wij adviseren op basis van deze bevinding het predictiemodel te gebruiken om na afronding van het oriënterend fertiliteitsonderzoek de spontane kans op zwangerschap te berekenen. Deze prognose kan waardevol zijn in de beslissing al dan niet te starten met fertiliteitsbehandeling.

- De voorspellende waarde van de obstetrische voorgeschiedenis om de kans op spontane zwangerschap te voorspellen berust voornamelijk op de manier van totstandkoming van de eerdere zwangerschap (spontaan of met ART), de lokalisatie van de eerdere zwangerschap (intra-uterien of extra-uterien) en of de eerdere zwangerschap ontstond in de huidige of een eerdere relatie.

- De uitkomst van de PCT kan in 50% van de subfertiele paren worden voorspeld, zonder dat de voorspellende waarde van de PCT zelf hiermee verloren gaat. Wij adviseren alleen bij die paren de PCT te verrichten, die primair subfertiel zijn en waarbij de man een normaal sperma heeft.

- De huidige WHO-criteria voor spermakwaliteit blijken slechts van geringe waarde te zijn in een grote populatie van subfertiele paren met betrekking tot het vermogen paren te identificeren die zwanger zullen worden zonder fertiliteitsbehandeling. De huidige WHO-criteria dienen te worden herzien door nieuwe criteria voor spermakwaliteit.

- Basaal FSH is een factor van belang bij de voorspelling van zwangerschapskansen. De rol van FSH in de beslissing om wel of niet te starten met fertiliteitsbehandeling is gering in een algemene subfertiele populatie. Wij adviseren om alleen basaal FSH te testen bij vrouwen tussen de 35 en 38 jaar, bij vrouwen met een korter wordende cyclus en vrouwen met een belaste familieanamnese voor prematuur ovarieel falen.

- Obesitas is een belangrijke risicofactor voor een afname van de zwangerschapskans van subfertiele vrouwen met een ovulatoire cyclus.

Authors' affiliations

PMM Bossuyt	Department of Clinical Epidemiology and Biostatistics, AMC, Amsterdam
PXJM Bouckaert	Department of Obstetrics & Gynaecology, Atrium Medisch Centrum, Heerlen
FJ Broekmans	Division of Reproductive Medicine, Department of Perinatology and Gynaecology, UMC, Utrecht
JM Burggraaff	Department of Obstetrics & Gynaecology, Scheperziekenhuis Emmen, Emmen
HJHM van Dessel	Department of Obstetrics & Gynaecology, TweeSteden Ziekenhuis, Tilburg
MJC Eijkemans	Department of Public Health, Erasmus MC, University MC, Rotterdam
PFM van der Heijden	Department of Obstetrics & Gynaecology, Twenteborg Ziekenhuis, Almelo
JDF Habbema	Department of Public Health, Erasmus MC, University MC, Rotterdam
PGA Hompes	Department of Obstetrics & Gynaecology, Vrije Universiteit MC, Amsterdam
JAM Kremer	Department of Obstetrics & Gynaecology, UMC St. Radbout, Nijmegen
L van der Leeuw-Harmsen	Department of Obstetrics & Gynaecology, Deventer Ziekenhuis, Deventer
HWB Michgelsen	Department of Obstetrics & Gynaecology, Medisch Spectrum Twente, Enschede
BWJ Mol	Center for Reproductive Medicine, AMC, Amsterdam Department of Obstetrics & Gynaecology, Máxima Medical Center, Veldhoven
GJE Oosterhuis	Department of Obstetrics & Gynaecology, Medisch Spectrum Twente, Enschede
JW van der Steeg	Department of Public Health, Erasmus MC, University MC, Rotterdam Center for Reproductive Medicine, AMC, Amsterdam Department of Obstetrics & Gynaecology, Vrije Universiteit MC, Amsterdam

P Steures	Department of Obstetrics & Gynaecology, Vrije Universiteit MC, Amsterdam Center for Reproductive Medicine, AMC, Amsterdam Department of Public Health, Erasmus MC, University MC, Rotterdam
F van der Veen	Center for Reproductive Medicine, AMC, Amsterdam

Members of the CECERM study Group

(Collaborative Effort for Clinical Evaluation in Reproductive Medicine)

Y.M. van Kasteren; Alkmaar, Medisch Centrum Alkmaar
P.F.M. van der Heijden; Almelo, Twenteborg Ziekenhuis
W.A. Schöls; Amersfoort, Meander Medisch Centrum
M.H. Mochtar; Amsterdam, Academisch Medisch Centrum
G.L.M. Lips; Amsterdam, BovenIJ Ziekenhuis
J. Dawson; Amsterdam, Sint Lucas Andreas Ziekenhuis
H.R. Verhoeve; Amsterdam, Onze Lieve Vrouwe Gasthuis
S. Milosavljevic; Amsterdam, Slotervaart Ziekenhuis
P.G.A. Hompes; Amsterdam, Vrij Universiteit Medisch Centrum
L.J. van Dam; Apeldoorn, Gelre Ziekenhuis
A.V. Sluijmer; Assen, Wilhelmina Ziekenhuis
H.E. Bobeck; Beverwijk, Rode Kruis Ziekenhuis
R.E. Bernardus; Blaricum, Tergooiziekenhuizen
M.C.S. Vermeer; Breda, Amphia Ziekenhuis Breda
J.P. Dörr; Den Haag, Medisch Centrum Haaglanden, locatie Westeinde
P.J.Q. van der Linden; Deventer, Ziekenhuis Deventer
H.J.M. Roelofs; Dordrecht, Albert Schweitzer Ziekenhuis
J.M. Burggraaff; Emmen, Scheper Ziekenhuis
G.J.E. Oosterhuis; Enschede, Medisch Spectrum Twente
M.H. Schouwink; Geldrop, Sint Anna Ziekenhuis
P.X.J.M. Bouckaert; Heerlen, Atrium Medisch Centrum
F.M.C. Delemarre; Helmond, Elkerliek Ziekenhuis
C.J.C.M. Hamilton; ’s Hertogenbosch, Jeroen Bosch Ziekenhuis
M. van Hoven; Hilversum, Tergooiziekenhuizen
M.H. Emanuel; Hoofddorp, Spaarne Ziekenhuis
C.N.M. Renckens; Hoorn, Westfries Gasthuis
J.A. Land; Maastricht, Academisch Ziekenhuis Maastricht
J.H. Schagen-Van Leeuwen; Nieuwegein, Sint Antonius Ziekenhuis
J.A.M. Kremer; Nijmegen, UMC Sint Radboud
C. van Katwijk; Purmerend, Waterland Ziekenhuis
M.H.A. van Hooff; Rotterdam, Sint Franciscus Gasthuis
H.J.H.M. Van Dessel; Tilburg/Waalwijk, TweeSteden Ziekenhuis
F.J.M. Broekmans; Utrecht, UMC Utrecht
H.J.L.A. Ruis; Veghel, Ziekenhuis Bernhoven
C.A.M. Koks; Veldhoven, Máxima Medisch Centrum

P. Bourdrez; Venlo/Venray, VieCuri Medisch Centrum
W.W.J. Riedijk; Zaandam, Zaans Medisch Centrum
B.J. Cohlen; Zwolle, Isala Klinieken

Publications by the author

Van der Steeg JW, Steures P, and Mol BWJ. (2003) Certainties and uncertainties in the basic fertility work-up. *Tijdschrift voor Fertiliteits Onderzoek* 17,42-46.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Van der Veen F, Bossuyt PMM, Hompes PGA, and Mol BWJ. (2004) Should the post-coital test (PCT) be part of the routine fertility work-up? *Hum Reprod* 19,1373-1379.

Van der Steeg JW, Steures P, Hompes PGA, Eijkemans MJC, Van der Veen F, and Mol BWJ. (2005) Investigation of the infertile couple: a basic fertility work-up performed within 12 months of trying to conceive generates costs and complications for no particular benefit. *Hum Reprod* 20,2672-2674.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Bossuyt PMM, Hompes PGA, Van der Veen F, and Mol BWJ. (2006) Which factors play a role in clinical decision-making in subfertility? *Reprod Biomed Online* 12,473-480.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JDF, Bossuyt PMM, Hompes PGA, Van der Veen F, and Mol BWJ. (2006) Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? *BJOG* 113,825-831.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Bouckaert PX, Bossuyt PM, Van der Veen F, and Mol BW. (2007) Predictive value and clinical impact of basal FSH in subfertile, ovulatory women. *The J Clin Endocrinol & Metab*.

Van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, Van Dessel HJ, Bossuyt PM, Van der Veen F, and Mol BWJ. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,536-542.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Burggraaff JM, Oosterhuis GJE, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2008) Obesity affects spontaneous pregnancy chances in subfertile, ovulatory women. *Hum Reprod* 23,324-328.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Michgelsen HWB, Van der Heijden PFM, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2007) Predictive value of pregnancy history in subfertile couples: results from a nationwide cohort study in the Netherlands. *Fertil Steril in press*

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Kremer JAM, Van der Leeuw-Harmsen L, Bossuyt PMM, Mol BWJ, and Van der Veen F. Evaluation of the WHO criteria for semen quality in a subfertile population: the need for new criteria. *Submitted*

Bongers MY, Bourdrez P, **Van der Steeg JW**, Heintz APM, Brolmann HAM, and Mol BWJ. (2005) Thermal endometrial ablation in dysfunctional uterine bleeding: An economic comparison of bipolar ablation and balloon ablation. *Gynecological Surgery* 2,271-277.

Coppus SFPJ, Verhoeve HR, Opmeer BC, **Van der Steeg JW**, Steures P, Eijkemans MJC, Hompes PGA, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2007) Identifying subfertile ovulatory women for timely tubal patency testing: a clinical decision rule based on medical history. *Hum Reprod* 2007, 22,2685-2692.

Coppus SFPJ, Land JA, Opmeer BC, Steures P, Eijkemans MJC, Hompes PGA, Bossuyt PMM, Van der Veen F, Mol BWJ, and **Van der Steeg JW**. (2007) Is positive *Chlamydia trachomatis* IgG serology associated with decreased treatment independent pregnancy rates in ovulatory subfertile women without tubal pathology? Submitted (*Hum Reprod*)

Cox T, **Van der Steeg JW**, Steures P, Hompes PGA, Van der Veen F, Eijkemans MJC, Schagen Van Leeuwen JH, Renckens C, Bossuyt PPM, Mol BWJ. (2008) Time to pregnancy after a previous miscarriage in subfertile couples. Submitted (*Hum Reprod*)

Custers IM, Steures P, **Van der Steeg JW**, Van Dessel HJHM, Bernardus RE, Bourdrez P, Koks CAM, Riedijk WJ, Burggraaff JM, Van der Veen F, and

Mol BWJ. (2007) External validation of a prediction model for an ongoing pregnancy after intrauterine insemination. *Fertility and Sterility*, 88, 425-431.

Graziosi GC, **Van der Steeg JW**, Reuwer PH, Drogtrop AP, Bruinse HW, and Mol BWJ. (2004) Economic evaluation of misoprostol in the treatment of early pregnancy failure compared to curettage after an expectant management. *Hum Reprod*. 20,1067-1071.

Leushuis E, **Van der Steeg JW**, Steures P, Repping S, Schöls W, Van der Veen F, Mol BWJ, and Hompes PGA (2008) Immunoglobulin G antisperm antibodies and prediction of spontaneous pregnancy. Accepted (*Fertil Steril*)

Steures P, **Van der Steeg JW**, Mol BWJ, Eijkemans MJC, Van der Veen F, Habbema JDF, Hompes PGA, Bossuyt PMM, Verhoeve HR, Van Kasteren YM, Van Dop PA. (2004) Prediction of an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 82,45-51.

Steures P, **Van der Steeg JW**, Verhoeve HR, Van Dop PA, Hompes PGA, Bossuyt PMM, Van der Veen F, Habbema JDF, Eijkemans MJC, and Mol BWJ. (2004) Does ovarian hyperstimulation in intrauterine insemination for cervical factor subfertility improve pregnancy rates? *Hum Reprod* 19,2263-2266.

Steures P, Berkhout JC, Hompes PGA, **Van der Steeg JW**, Bossuyt PMM, Van der Veen F, Habbema JDF, Eijkemans MJC, and Mol BWJ. (2005) Patients' preferences in deciding between intrauterine insemination and expectant management. *Hum Reprod* 20,752-755.

Steures P, **Van der Steeg JW**, Hompes PGA, Habbema JDF, Eijkemans MJC, Broekmans FJ, Verhoeve HR, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2006) Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet* 368,216-221.

Steures P, **Van der Steeg JW**, Hompes PGA, Van der Veen F, and Mol BWJ. (2006) [Results of intrauterine insemination in the Netherlands]. *Ned Tijdschr Geneeskd* 150,1127-1133.

Steures P, **Van der Steeg JW**, Hompes PGA, Van der Veen F, and Mol BWJ. (2007) Intrauterine insemination in The Netherlands. *Reprod Biomed Online* 14,110-116.

Steures P, **Van der Steeg JW**, Hompes PGA, Bossuyt PMM, Habbema JDF, Eijkemans MJC, Koks CAM, Boudrez P, Van der Veen F, and Mol BW. (2007) The additional value of ovarian hyperstimulation in intrauterine insemination for couples with an abnormal postcoital test and a poor prognosis: a randomized clinical trial. *Fertil Steril* 88,1618-1624.

Steures P, **Van der Steeg JW**, Hompes PGA, Bossuyt PMM, Habbema JDF, Eijkemans MJC, Schöls WA, Burggraaff JM, Van der Veen F, and Mol BWJ. (2007) Effectiveness of intrauterine insemination in subfertile couples with an isolated cervical factor: a randomized clinical trial. *Fertil Steril* 88,1692-1696.

Steures P, **Van der Steeg JW**, Hompes PGA, Bossuyt PMM, Mol BWJ, Van der Veen F (2008) Intrauterine insemination, what do we really know? A critical appraisal of the literature. Submitted (*Human Rep Update*)

Steures P, **Van der Steeg JW**, Hompes PGA, Habbema JDF, Eijkemans MJC, Broekmans FJ, Verhoeve HR, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2008) [Intrauteriene inseminatie met gecontroleerde ovariële hyperstimulatie in vergelijking met een afwachtend beleid bij paren met onverklaarde subfertiliteit en een gemiddelde prognose: een gerandomiseerd onderzoek. Accepted (*Ned Tijdschr Geneeskd)*

Van Weert JM, Repping S, **Van der Steeg JW**, Steures P, Van der Veen F, and Mol BWJ. (2005) IUI in male subfertility: are we able to select the proper patients? *Reprod Biomed Online* 11,624-631.

Van Weert JM, Van den Broek J, **Van der Steeg JW**, Van der Veen F, Flierman PA, Mol BWJ, and Steures P. (2007) Patients' preferences for intrauterine insemination or in vitro fertilization. *RBM online*,15, 422-427.

Vellekoop J, Vrouenraets FPJM, **Van der Steeg JW**, Mol BWJ, Roumen FJME. (2008) Factors associated with the decision to induce labor: a survey among Dutch gynaecologists and midwives. Submitted (*BJOG*)

Letters to the editor

Steures P, Hompes PGA, **Van der Steeg JW**, Eijkemans MJC, and Mol BWJ. (2003) Life-table analysis falls short of the mark! *Fertil Steril* 79,1467-1468.

Van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Broekmans FJ, Van Dessel HJHM, Bossuyt PMM, Van der Veen F, and Mol BWJ. (2007) Reply: Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod* 22,2345-2346.

Van der Steeg JW, Steures P, Te Velde ER, Hompes PGA, and Mol BWJ Treatment strategies for subfertile couples. (2004) *Hum Reprod* 19,1678.

Dankwoord

Eindelijk is het zover, het proefschrift waar de afgelopen jaren zo hard aan is gewerkt, ligt voor u! Het spreekt voor zich dat hier de bijdrage van zeer veel mensen onmisbaar is geweest. Om uiteenlopende redenen ben ik de volgende mensen dankbaar.

Om te beginnen, Pieternel, jij bent in dit onderzoeksproject echt een maatje door dik en dun geweest. Letterlijk, wat betreft de vele afslankdrankjes gecombineerd met overheerlijke brownies… Figuurlijk, vanwege de vele bijzondere momenten die we hebben meegemaakt, wetenschappelijke hoogtepunten zoals het congres in Montréal en het Lancet stuk, maar ook moeilijke, soms emotionele momenten, waar we altijd weer veel steun aan elkaar hadden. De OFO-feestweek is onze welverdiende beloning!

De promotoren, prof. dr. J.D.F. Habbema en prof. dr. F. van der Veen dank ik voor de mogelijkheid die ze mij hebben gegeven dit onderzoek te doen. Dik, het vertrouwen dat je me hebt geven tijdens dit onderzoek, kreeg ik al bij de eerste kop koffie op Utrecht Centraal 2002; waar nodig heb je me vrij gelaten het onderzoek naar eigen inzicht uit te voeren. Fulco, wat is het leuk om via een ander te horen dat je trots op me bent. Het enthousiasme dat je uitstraalt als jij in iets gelooft dat we net samen bedacht hebben, is onbeschrijfelijk. En nu nog samen bewijzen dat infertiliteit echt niet aan de man ligt!

De co-promotoren, prof. dr. B.W.J. Mol en dr. M.J.C. Eijkemans. Ben Willem, wat een onuitputtelijke hoeveelheid ideeën en energie heb jij toch altijd. Bij het per e-mail versturen van een stuk, leek het vaak net of ik auto-reply had geactiveerd, echter dat was meestal al jouw antwoord met commentaar! Ik wens je veel geluk met je welverdiende hoogleraarschap. René, jouw geduldige manier van statistiek uitleggen, samen analyses doen en uiteindelijk het mij zelf laten uitvoeren is uniek. Ik heb heel veel van je geleerd. De appelflappen tussen de analyses door niet te vergeten. Ik hoop je nog eens te mogen bellen voor lastige berekeningen.

De overige OFO-projectleden prof. dr. P.M.M. Bossuyt en dr. P.G.A. Hompes. Beste Patrick, als ik ergens echt niet meer uitkwam, kon ik altijd, op advies van Ben Willem, bij jou terecht. Ik ben nooit weggegaan zonder een oplossing. Peter, as ever, dank zij onder andere jouw initiatief met de FC

Noord Holland heb ik dit onderzoek kunnen doen. Zelfs als een stuk acht keer is afgewezen, bleef je positief!

Research nurse Karin van Schaik. Karin, geweldig dat je mij en Pieternel zo hebt bijgestaan in het OFO-project. Met jouw nieuwsbrieven en OFO-presentjes heeft het onderzoek echt glans gekregen.
De commissie prof. dr. E.A.P. Steegers, prof.dr. G.J. Bonsel, prof.dr. B.C.J.M. Fauser, prof.dr. E.W. Steijerberg, prof.dr. J.A.M. Kremer, bedankt voor het beoordelen van het manuscript.

De patiënten wil ik bedanken, die in hun onzekere tijd rondom de onvervulde kinderwens toch bereid waren deel te nemen aan het onderzoek.

Ik dank alle gynaecologen van het OFO-project voor hun inzet: Y.M. van Kasteren, P.F.M. van der Heijden, W.A. Schöls, M.H. Mochtar, G.L.M. Lips, J. Dawson, H.R. Verhoeve, S. Milosavljevic, P.G.A. Hompes, L.J. van Dam, A.V. Sluijmer, H.E. Bobeck, R.E. Bernardus, M.C.S. Vermeer, J.P. Dörr, P.J.Q. van der Linden, H.J.M. Roelofs, J.M. Burggraaff, G.J.E. Oosterhuis, M.H. Schouwink, P.X.J.M. Bouckaert, F.M.C. Delemarre, C.J.C.M. Hamilton, M. van Hoven, M.H. Emanuel, C.N.M. Renckens, J.A. Land, J.H. Schagen-Van Leeuwen, J.A.M. Kremer, C. van Katwijk, M.H.A. van Hooff, H.J.H.M. Van Dessel, F.J.M. Broekmans, H.J.L.A. Ruis, C.A.M. Koks.

Daarnaast wil ik alle fertiliteitsartsen, poliverpleegkundigen en doktersassistenten bedanken voor het patiëntencontact, invoerwerk en het verzamelen van dossiers.

Medewerkers Erasmus MC: Wilma, Esther, Judith, Merel, Cecile en Ewout en alle andere collega's; ondanks dat ik er niet vaak was, gaven jullie mij het gevoel dat ik erbij hoorde.

Medeonderzoekers AMC: Wouter, Janne-Meije, Inge, Esther, Judith, Wessel, Liesbeth, Bas, Wouter, Moira, Saskia, Marja, Madelon, Neriman, Henrike, Maureen, Marjolein, Karlijn, Christianne, Moniek, Stef, Jiska, Maarten, Etelka, Nicole en Sjors; de vele onderzoekerslunches, de Douwe Egberts koffiemanager en de 2x2 tabellen hebben ons erdoor gesleept. Alle nieuwe onderzoekers wens ik succes en evenveel onderzoekersplezier!

Alle collega-assistenten, stafleden, verpleegkundigen, polidames, jullie wil ik danken voor je interesse en hulp bij de voorbereidingen van de promotie.

Alle collega's uit het St. Lucas Andreas Ziekenhuis; dank voor jullie interesse, maar ook voor het vertrouwen en het gevoel er helemaal bij te horen tijdens mijn eerste opleidingsdeel.

Mirelle, jouw artistieke kijk op mijn onderzoek heeft dit prachtige boek opgeleverd.

Annemarie, dank voor je hulp en geduld bij het maken van mijn proefschrift. Ik denk inderdaad dat jij van iedereen mijn boek het vaakst hebt doorgelezen.

Joost, Heyn, Remko, Bob, bedankt voor de gezellige kookavonden, die voor de o zo belangrijke ontspanning zorgden. Alle andere vrienden, bedankt voor jullie interesse.

Lieve oma Hamoen, dit boek draag ik ook een beetje aan u op. Ik vind het ongelooflijk dat iemand op 97e jarige leeftijd nog precies begrijpt waar mijn onderzoek over gaat. U bent voor mij een heel bijzonder mens.

Lieve Jos en Corry, bedankt voor jullie steun en liefde.

Lieve Jan en Thea, wat ben ik trots dat jullie mijn ouders zijn. Papa, ik heb zoveel van je geleerd, elke dag als ik met mensen omga, of als ik opereer, komt dat weer van pas. Wat ben ik blij dat je weer de energie uitstraalt zoals ik die van je ken. Ik hoop ooit nog eens van je te winnen met hardlopen… Mama, zo zorgzaam en geïnteresseerd als jij is er maar een!

Beste Martijn, een jaar na jouw promotie mag je dan mijn paranimf zijn. Super om sinds Antwerpen je maatje te zijn. Op naar een weekend Leicester!

Bruur Rolf, dat jij als paranimf mij terzijde staat, is voor mij een hele eer. Sorry dat je als paranimf net wat meer moet doen dan als getuige…

Lieve Marieke, je zou het afschuwelijk vinden als ik hier schreef wat je allemaal voor mij betekent. Dat zal ik dan ook niet doen. Lieverd, ik hou van je. Lieve Janneke en Floor, wat zijn jullie toch een prachtige dochters, van wie zouden jullie dat eigenwijze karakter toch hebben… Heerlijk om jullie na een dag hard werken in mijn armen te hebben. Lieve Wouter, als iemand weet van doorzetten, dan bij jij het wel! Je hebt letterlijk mijn hart gestolen!