Human Reproduction Update Advance Access published May 12, 2009

Human Reproduction Update, Vol.1, No.1 pp. 1–16, 2009

doi:10.1093/humupd/dmp013

human reproduction update

Prediction models in reproductive medicine: a critical appraisal[†]

Esther Leushuis[1,2,5], Jan Willem van der Steeg[2], Pieternel Steures[2], Patrick M.M. Bossuyt[3], Marinus J.C. Eijkemans[4], Fulco van der Veen[2], Ben W.J. Mol[2,3], and Peter G.A. Hompes[1]

[1]Department of Obstetrics and Gynecology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands [2]Department of Obstetrics/Gynaecology, Center for Reproductive Medicine, Academic Medical Center, Room H4-238, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands [3]Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands [4]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

[5]Correspondence address. Tel: +31-20-5663456; Fax: +31-20-6963489; E-mail: e.leushuis@amc.uva.nl

TABLE OF CONTENTS

BACKGROUND: Prediction models have been developed in reproductive medicine to help assess the chances of a treatment-(in)dependent pregnancy. Careful evaluation is needed before these models can be implemented in clinical practice.

METHODS: We systematically searched the literature for papers reporting prediction models in reproductive medicine for three strategies: expectant management, intrauterine insemination (IUI) or *in vitro* fertilization (IVF). We evaluated which phases of development these models had passed, distinguishing between (i) model derivation, (ii) internal and/or external validation, and (iii) impact analysis. We summarized their performance at external validation in terms of discrimination and calibration.

RESULTS: We identified 36 papers reporting on 29 prediction models. There were 9 models for the prediction of treatment-independent pregnancy, 3 for the prediction of pregnancy after IUI and 17 for the prediction of pregnancy after IVF. All of the models had completed the phase of model derivation. For six models, the validity of the model was assessed only in the population in which it was developed (internal validation). For eight models, the validity was assessed in populations other than the one in which the model was developed (external validation), and only three of these showed good performance. One model had reached the phase of impact analysis.

CONCLUSIONS: Currently, there are three models with good predictive performance. These models can be used reliably as a guide for making decisions about fertility treatment, in patients similar to the development population. The effects of using these models in patient care have to be further investigated.

Key words: prediction model / fertility / pregnancy / spontaneous pregnancy / ART (IUI/IVF)

Introduction

Until recently, the emphasis in reproductive medicine has been on finding causal diagnoses of subfertility followed by treatment of the diagnosed condition. Examples are ovulation induction in women diagnosed with anovulation, tubal surgery in women with bilateral tubal disease and *in vitro* fertilization (IVF) with assisted fertilization after surgical sperm retrieval in couples with azoospermia. In many couples, such causal factors cannot be found. These couples are classified as having unexplained subfertility, mild male subfertility, cervical factor subfertility, mild endometriosis or one-sided tubal pathology; and assisted reproductive techniques such as intrauterine insemination

[†]Abstract presented at the 24th Annual Meeting of the European Society of Human Reproduction and Embryology (ESHRE) in Barcelona, Spain (6–9 July 2008).

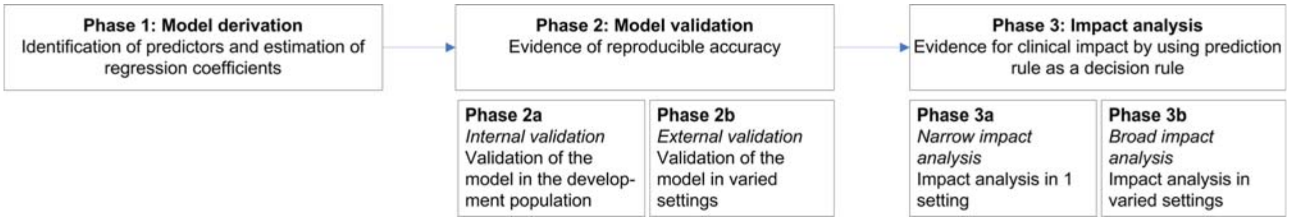
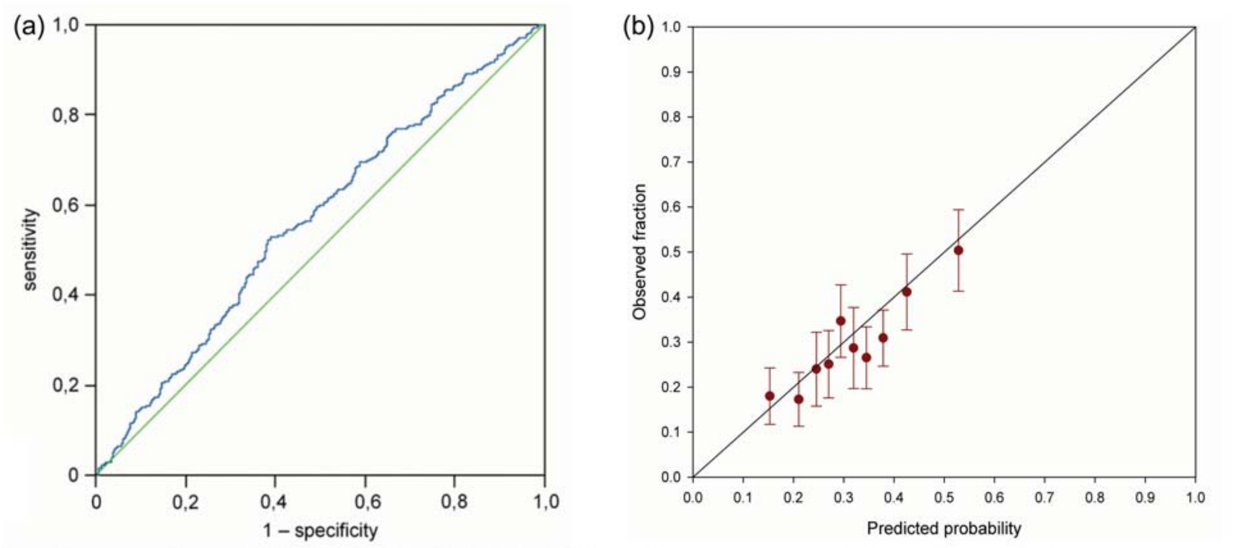

Phase 1: Model derivation
Identification of predictors and estimation of regression coefficients
Phase 2: Model validation
Evidence of reproducible accuracy
Phase 3: Impact analysis
Evidence for clinical impact by using prediction rule as a decision rule
Phase 2a
Internal validation
Validation of the model in the development population
Phase 2b
External validation
Validation of the model in varied settings
Phase 3a
Narrow impact analysis
Impact analysis in 1 setting
Phase 3b
Broad impact analysis
Impact analysis in varied settings

Figure 1 Phases of model development.



From: Custers *et al.* External validation of model for IUI. Fertil Steril 2007.

From: van der Steeg et *al.*. Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. Human Reproduction 2007.

Figure 2 (**a**) Typical ROC curve of a prediction model in reproductive medicine (AUC 0.56) and (**b**) calibration plot with calculated probability on the *X*-axis and observed proportion on *Y*-axis, showing good calibration.

(IUI) or IVF are then considered. As these interventions are expensive and not without side effects, they should be offered to a couple only if the expected success rate with treatment substantially exceeds the probability of a spontaneous pregnancy (Wasson *et al.*, 1985; te Velde and Cohlen, 1999).

It is a clinical challenge for gynaecologists to make such a comparison. Gynaecologists are known to differ widely in their estimations of the probability of achieving a pregnancy for subfertile couples (van der Steeg *et al.*, 2006). To help gynaecologists in assessing the chances of pregnancy, prediction models have been developed. With these models, one can calculate the probability of a treatment-independent pregnancy as well as the probability of success with IUI and IVF.

Careful evaluation is needed before these models can be implemented in clinical practice. The use of poor-quality prediction models could have a negative effect on decision making, by introducing the illusion of objective improvement over clinical judgment. We systematically reviewed the literature on the available prediction models in reproductive medicine. We appraised the prediction models according to a published evaluation scheme, distinguishing between model derivation, model validation and impact analysis (McGinn *et al.*, 2000; Reilly and Evans 2006; Steyerberg, 2008). We also summarized their performance.

Methods

Search strategy and study selection

We performed a structured predefined literature search using MEDLINE, EMBASE and the Cochrane Library from inception to October 2008. An information specialist performed the electronic search using the following terms: pregnancy, live birth, conception, infertility/subfertility/fertility, intrauterine insemination, *in vitro* fertilization, prediction models and validation. We checked cross-references of eligible papers to identify other papers not captured by electronic searches. No restrictions were held concerning publication year or language. A Reference Manager 11.0 database was established to incorporate results of all citations.

Two reviewers (J.W.S. and E.L.) evaluated potentially eligible papers in a two-stage process. First, papers identified in the search were independently screened for eligibility by reading the title and abstract. If there were any doubts about eligibility after reading the title, we screened the

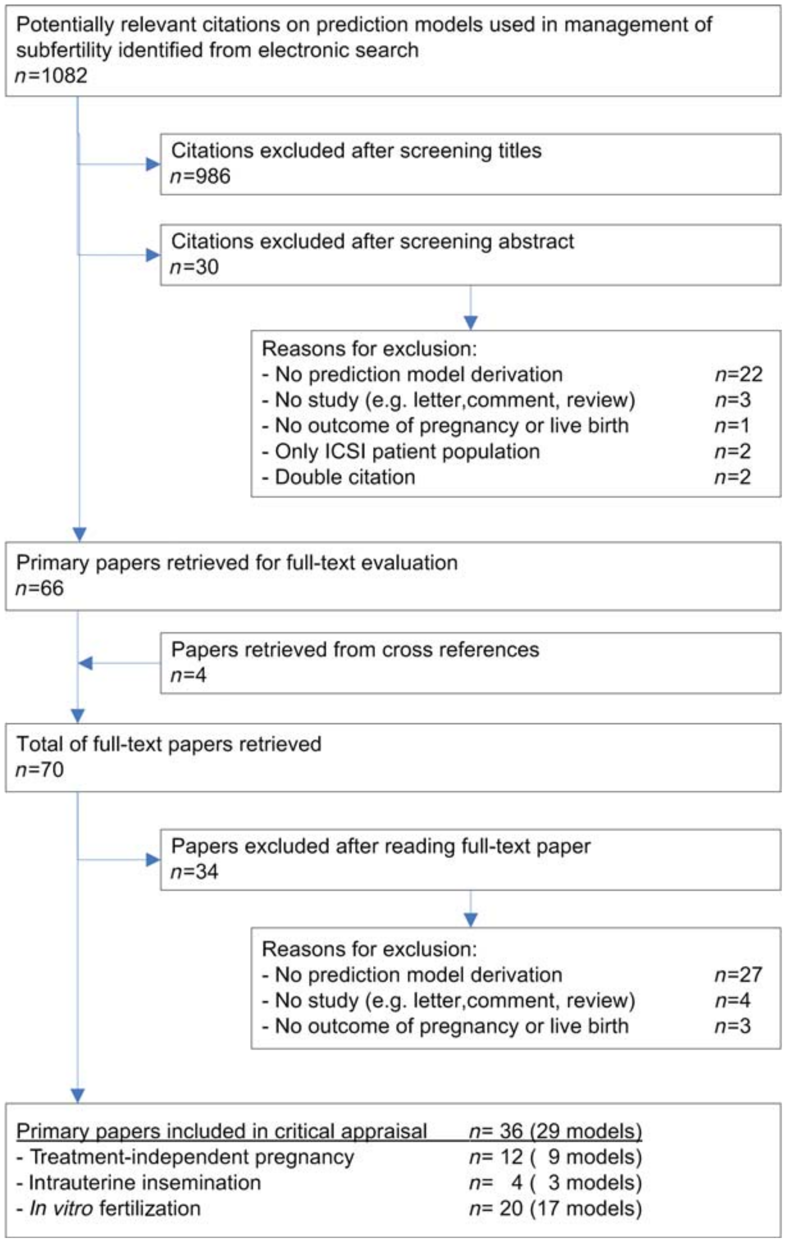

Potentially relevant citations on prediction models used in management of subfertility identified from electronic search
n=1082
Citations excluded after screening titles
n=986
Citations excluded after screening abstract
n=30
Reasons for exclusion:
- No prediction model derivation n=22
- No study (e.g. letter,comment, review) n=3
- No outcome of pregnancy or live birth n=1
- Only ICSI patient population n=2
- Double citation n=2
Primary papers retrieved for full-text evaluation
n=66
Papers retrieved from cross references
n=4
Total of full-text papers retrieved
n=70
Papers excluded after reading full-text paper
n=34
Reasons for exclusion:
- No prediction model derivation n=27
- No study (e.g. letter,comment, review) n=4
- No outcome of pregnancy or live birth n=3
Primary papers included in critical appraisal n= 36 (29 models)
- Treatment-independent pregnancy n= 12 (9 models)
- Intrauterine insemination n= 4 (3 models)
- In vitro fertilization n= 20 (17 models)

Figure 3 Process from initial search to final inclusion for papers on prediction models in reproductive medicine.

abstract and the full text to make sure no papers were missed. We then obtained full-text versions of all papers selected by at least one of the reviewers in the first stage. Papers were included if they reported on a prediction model for treatment-independent pregnancy, pregnancy after IUI or IVF. If the paper reported on a model for embryo transfer only, it was excluded.

In this review, a prediction model was defined as a model that expressed pregnancy as a function of one or more predictor variables. Such a model can be based on a multivariable regression model, such as a linear, logistic or Cox proportional hazards regressions model. To be eligible, the reported prediction model had to be presented as a score chart, a prediction rule or as a set of regression coefficients with baseline intercept, sufficient to make predictions for individual cases.

Assessment of study quality

For each included paper, we identified the study characteristics and assessed the study quality on the basis of the following items for all models: whether the patient selection was consecutive, whether the data had been collected prospectively, whether the variables and pregnancy (or live birth) were described in sufficient detail and whether missing data were reported and/or imputed (‘filled in’). For papers that reported on treatment-independent pregnancy, the basic fertility work-up had to be clearly described, no treatment between basic fertility work-up and time to pregnancy should have been applied and the follow-up duration had to be at least 1 year. We also verified whether papers that reported on treatment-independent models had been derived from Cox proportional hazards analysis with or without right-hand

Table I Study characteristics of the papers that report on prediction models for treatment-independent pregnancy

First author (year)	Patients	Inclusion and exclusion criteria	*n*	Study design[b]	Outcome[c]
Jedrzejczak *et al.* (2008)	Men from infertile couples without female infertility factor (cases) matched with healthy fertile sperm donors (controls)	Exclusion (cases): - azoospermia - total lack of sperm motility Exclusion (controls): - history of past infertility - history of inflammation or surgery of the reproductive organs <1 year	242	cc study	preg.
van der Steeg *et al.* (2007)	Subfertile couples not evaluated previously for subfertility referred by a general practitioner External validation of Hunault *et al.* (2004)	Exclusion: - ovulation disorder - TMC $<3\times10^6$ - one- or two-sided tubal pathology	3021	pros. CH	ong. preg.
Hunault (2005)	Couples from two university hospitals with subfertility due to mild male, cervical or unexplained subfertility External validation of Hunault *et al.* (2004)	Inclusion: - woman's age <40 years Exclusion: - ovulation disorder - azoospermia - one-sided/two-sided tubal defect - endocrine disorder	302	pros. CH	live birth
Hunault *et al.* (2004)	Patients from Snick *et al.* (1997), Collins *et al.* (1995) and Eimers *et al.* 1994[a] External validation of Snick *et al.* (1997), Collins *et al.* (1995) and Eimers *et al.* (1994)	Exclusion: - ovulation disorder - tubal pathology - azoopermia	[a]	pros. CH	live birth
Hunault (2002a)	First visit of subfertile couples at an university fertility clinic External validation of alternate model of Eimers *et al.* (1994)	Exclusion: - ovulation disorder - azoospermia - one-sided/two-sided tubal defect	1061	pros. CH	live birth
Snick *et al.* (1997)	Subfertile couples from a secondary care fertility centre	Inclusion: - child wish - >1 year no pregnancy	726	pros. CH	live birth
Collins *et al.* (1995)	First visit of subfertile couples at an university fertility clinic	No exclusion criteria reported	2198	pros. CH	live birth
Bahamondes *et al.* (1994)	Subfertile couples consulting infertility clinic with 3 years of follow-up or pregnancy	Exclusion: - divorced during study - history of tubal ligation or habitual abortion - azoospermia	559	ret. CH	preg.
Wichmann *et al.* (1994)	Subfertile men, referred to andrological laboratory for subfertility problems with registered duration of subfertility	Exclusion: – abstinence period <3 days - incomplete sample - azoospermia - donor insemination	907	pros. CH	preg.
Eimers *et al.* (1994)	Subfertile couples from a university fertility centre	Inclusion: - cycle of 23–35 days - biphasic BTC - no azoospermia - no abnormal HSG or laparoscopy	996	Pros. CH	preg.
Bostofte *et al.* (1993)	Subfertile couples investigated for subfertility at a university hospital	No exclusion criteria reported	321	pros. CH	preg.
Bostofte (1987)	Men with semen analysis for subfertility who responded to a questionnaire	Exclusion: - azoospermia - invalid name and birth - death/emigration - not traceable in official registers	765	ret. CH	preg.

[a] For details, see Snick *et al.* (1997), Collins *et al.* (1995) and Eimers *et al.* (1994).

[b] Study design: cc study = case control study; pros. CH = prospective cohort study; ret. CH = retrospective cohort study.

[c] Outcome: preg. = pregnancy; ong.preg. = ongoing pregnancy.

Table II Study characteristics of the papers that report on prediction models for pregnancy after IUI

First author (year)	Patients	Inclusion and exclusion criteria	*n*	Study design[a]	Outcome[b]
Erdem *et al*. (2008)	Patients with unexplained, mild male infertility with regular menstrual cycles	Inclusion: - midluteal progesterone $>$10 ng/ml - confirmed bilateral tubal patency - normal semen analysis (WHO, 1992) Exclusion: - PCOS - previous ovarian surgery - total motile sperm count (TMC) $<1 \times 10^6$/ml post-wash	456	ret. CH	live birth
Custers *et al*. (2007)	Couples treated with IUI External validation of Steures *et al*. (2004)	Inclusion: - confirmed ovulatory cycle - at least one patent tube	1079	pros. CH	ong.preg.
Steures *et al*. (2004)	Couples treated with IUI	Inclusion: all women with IUI for reasons of: - male factor - cervical factor - unexplained subfertility	3371	ret. CH	live birth
Tomlinson *et al*. (1996)	Couples treated with IUI	Inclusion: all women with IUI for reasons of: - unexplained subfertility - mild sperm dysfunction - anovulation - cervical mucus hostility	260	ret. CH	preg.

[a]Study design: pros. CH = prospective cohort study; ret. CH = retrospective cohort study.
[b]Outcome: preg. = pregnancy; ong.preg. = ongoing pregnancy.

censoring. We added two items for papers that reported on the prediction of treatment-dependent pregnancy (IUI or IVF): whether the diagnosis before treatment was described in sufficient detail and whether the protocol of treatment was described in sufficient detail. We report study quality separately for treatment-independent and treatment-dependent models, i.e. IUI and IVF.

Assessment of model development

We assessed the development of the prediction models with a published evaluation scheme, which distinguishes three phases: model derivation, model validation and impact analysis (Fig. 1) (McGinn *et al*., 2000; Reilly and Evans 2006; Steyerberg, 2008). In the model derivation phase, predictors are identified, based on prior knowledge, and the weight of each predictor (regression coefficient) is calculated. In the second phase, one can distinguish between the internal validation (phase 2a) and the external validation (phase 2b). With internal validation of a prediction model, the model's ability to predict outcome in the group of patients in which it was developed is evaluated, sometimes with data collected in a separate group of patients evaluated in the same setting (Altman and Royston, 2000). As internal validation systematically gives a too optimistic impression about the quality of the predictions, external validation is a vital next step in assessing the performance of the model (Harrell *et al*., 1996). With external validition of a prediction model, the model's ability to predict outcome in populations other than the population in which the model was developed, also called 'generalizability' or 'transportability', is evaluated. The third and final phase consists of impact analysis, which is the evaluation of the implementation of prediction models with documented validity. Impact analysis establishes whether the prediction model improves doctors' decisions by evaluating the effect on patient outcome (Reilly and Evans 2006; Steyerberg, 2008). This can be evaluated in one (phase 3a) or in varied settings (phase 3b), preferably in a randomized controlled trial.

Assessment of model performance

For the models that were evaluated in an external validation, we quantified the performance of the prediction models by assessing discrimination and calibration. Discrimination refers to the ability to distinguish couples who will conceive from those who will not. If there are multiple scores or probabilities, the sensitivity–specificity pairs for each cut-off value can be plotted in a receiver operating characteristic (ROC) curve (Fig. 2a) (Hanley and McNeil, 1982). In that case, discrimination can be expressed as the area under this ROC curve (AUC) or the c-statistic (Tosteson *et al*., 1994). An AUC of 1 implies perfect discrimination, whereas an AUC of 0.5 means that the test does not discriminate at all (Hanley and McNeil, 1982). For this review, a model is considered to have poor performance if the AUC lies between 0.50 and 0.70. An AUC between 0.70 and 0.80 represents fair performance, and an AUC between 0.80 and 0.90 represents good performance.

Calibration refers to the level of correspondence between the calculated pregnancy chances and the observed proportion of pregnancies. Calibration can be evaluated by several techniques of which we will describe the three techniques that are most commonly used. The first technique relies on a goodness-of-fit test for the model for predicting pregnancy (Hosmer, 2000). The second technique uses the coefficients of the linear regression line through the prediction–observation pairs in a calibration plot to evaluate the performance of a model. If the calibration is perfect, the line will be on the diagonal, with intercept zero and slope unity (Cox, 1958). For models with a slope below 1, high-probability predictions are too high and low-probability predictions are too low. If the slope exceeds 1, the bias is the other way around (Steyerberg *et al*.,

Table III Study characteristics of the papers that report on prediction models for pregnancy after IVF

First author (year)	Patients	Inclusion and exclusion criteria	*n*	Study design[d]	Outcome[e]
van Weert *et al*. (2008)	All couples with male subfertility undergoing IVF treatment	Inclusion: - 2 semen analyses that did not meet WHO criteria - oocyte retrieval	275 ptn[c]	ret. CH	ong.preg.
Hunault *et al*. (2007)	Patients from a university hospital in their first IVF cycle External validation of Hunault (2002b)	Inclusion:- transfer of two embryos Exclusion: - ICSI treatment - oocyte donation - cryopreserved embryos	642 ptn	ret. CH	ong.preg.
Lintsen *et al*. (2007)	Couples eligible for IVF and ICSI[a]	Exclusion: - no record of follow-up dates - no start of treatment for known reasons	4928 ptn	pros. CH	ong.preg.
Verberg *et al*. (2007)	Infertile patients with a regular indication for IVF or ICSI at an university hospital	Inclusion: - menstrual cycle 25–35 days - BMI 18–28 kg/m^2 Exclusion: - previous IVF - unhealthy child after IVF - frozen embryos transfer	201 ptn	pros. CH	ong.preg.
Carrera-Rotllan *et al*. (2007)	Patients with primary infertility due to a tubal factor with normal semen parameters at their first IVF attempt	Inclusion: - age <38 years - menstrual cycle 24–32 days - normal FSH/LH/E2/prolactin/TSH/BMI Exclusion: - age ≥ 38 years - history of genetic risks/pregnancy loss or preimplantation genetic diagnosis	110 ptn	pros. CH	preg.
Ottosen *et al*. (2007)	IVF and ICSI treatment cycles from a public fertility clinic	Exclusion: - frozen embryo replacement - single embryo transfer	2193 cyc.	ret. CH	preg.
Ferlitsch *et al*. (2004)	Women referred for IVF to a university hospital of known height and weight at their initial IVF cycle	Exclusion: - severe endometriosis - a single ovary with a possible normal ovarian response - any ovarian cyst measuring >10 mm in diameter on a baseline day	170 ptn	ret. CH	preg.
Hunault (2002b)	Women undergoing their first IVF cycle	Exclusion: - single embryo transfer (ET) - oocyte donation - cryothawed embryo cycles - ICSI - cycles not resulting in ET	642 ptn	ret. CH	ong.preg.
Smeenk *et al*. (2000)	Couples who started their first IVF cycle in a university hospital External validation of Templeton *et al*. (1996)	Exclusion: - ICSI cycles - donor gametes - frozen embryos	1253 ptn	pros. CH	ong.preg.
Stolwijk *et al*. (2000)	Couples who underwent their first IVF or ICSI treatment at a university fertility centre	Inclusion: - ≤41 yr and FSH <20 IU/L Exclusion: - donor semen - MESA or TESE[b] - donor oocytes	1315 ptn	pros. CH	ong.preg.

Continued

Table III *Continued*

First author (year)	Patients	Inclusion and exclusion criteria	*n*	Study design[d]	Outcome[e]
Bancsi *et al.* (2000)	Women undergoing their first stimulated IVF cycle at an academic fertility centre	Inclusion: - regular menstrual cycle - bFSH level on day 1–4 - no endocrine disorder - no oocyte donation - no unstimulated cycles	435 ptn	ret. CH	ong.preg.
Stolwijk *et al.* (1998)	Complete IVF cycles with hormone ovulation induction	Exclusion:	757 cyc.	pros. CH	ong.preg.
		- ICSI	432 cyc.		
	External validation of Stolwijk *et al.* (1996)	- donor oocytes	428 cyc.		
		- donor spermatozoa	1424 cyc.		
		- IVF for unexplained subfertility			
Minaretzis *et al.* (1998)	Consecutive IVF cycles	Inclusion: - at least one embryo transfer	544 ptn	pros. CH	live birth
Commenges-Ducos *et al.* (1998)	Consecutive IVF-embryo transfer cycles	Exclusion: - hyperandrogenism - uterine malformation - diethylstilboestrol syndrome - age ≥40 years with abnormal ovarian test reserve - cryo- or donation oocytes	923 cyc.	ret. CH	ong.preg.
Templeton *et al.* (1996)	All IVF treatment cylces in a national database	Exclusion: - sperm, oocyte or embryo donation - frozen embryo transfer - microassisted fertilization - unstimulated cycle (natural IVF)	36 961 cyc.	ret. CH	live birth
Stolwijk *et al.* (1996)	Couples who underwent their first IVF cycle	Exclusion: - donor oocytes - ICSI	757 cyc.	ret. CH	ong.preg.
Bouckaert *et al.* (1994)	Patients treated for IVF	No exclusion criteria reported	591 ptn	ret. CH	preg.
Haan *et al.* (1991)	All regular treatment cycles from five IVF centres	No exclusion criteria reported	3092 cyc.	pros. CH	ong.preg.
Hughes *et al.* (1989)	Consecutive IVF cycles	No exclusion criteria reported	716 cyc.	pros. CH	ong.preg.
Nayudu *et al.* (1989)	IVF patients with follicular aspirate	Exclusion: - ectopic pregnancy - post 13 weeks abortion - follicular fluid not present for technical reasons	222 ptn	ret. CH	ong.preg.

[a] ICSI = intracytoplasmatic sperm injection.
[b] MESA = microepididymal sperm aspiration; TESE = testicular sperm extraction.
[c] ptn = patients; cyc. = cycles.
[d] Study design: pros. CH = prospective cohort study; ret. CH = retrospective cohort study.
[e] Outcome: preg. = pregnancy; ong.preg. = ongoing pregnancy.

2001). A third technique for assessing the calibration is based on a visual interpretation of the calibration figure plot (Fig. 2b). A calibration plot is constructed by comparing the mean predicted probability (*X*-axis) with the observed proportion of pregnancies (*Y*-axis). For example, patients can be allocated to one of 10 groups of equal size on the basis of the deciles of the calculated probabilities. For each group, the mean predicted probability is calculated, as well as the observed proportion is calculated by Kaplan–Meier analysis. In case of perfect calibration, the prediction–observation pairs are on the main diagonal and confidence intervals are not overlapping. Points below the diagonal represent overestimation of the probability of pregnancy, and points above represent underestimation (Custers *et al.*, 2007; van der Steeg *et al.*, 2007). When impact analysis was performed, we evaluated the correspondence between the calculated probabilities and the observed percentage of pregnancies after the introduction of the prediction models.

Results

Our search retrieved 1082 citations from MEDLINE and EMBASE, and none from the Cochrane Library. The process of selection of papers is summarized in Fig. 3. We retrieved four papers from cross-references. After screening titles, abstracts and cross-references, we selected 70 papers for further reading. Exclusion criteria are shown in Fig. 3.

Table IV Overview of the parameters of the prediction models for treatment-independent pregnancy (expressed as HRs or ORs)

	Jedrzejczak *et al.* (2008)	Hunault *et al.* (2004)	Snick *et al.* (1997)	Collins *et al.* (1995)	Bahamondes *et al.* (1994)	Wichmann *et al.* (1994)	Eimers *et al.* (1994)	Bostofte *et al.* (1993)	Bostofte *et al.* (1987)	Presence of the parameter in the prediction model (number out of 9 models)
Type of analysis	LR	CR	CR	CR	LR	CR	CR	CR	CR	
Couple factors										
Duration of subfertility (year)		0.83	1.49[b]	1.68[c]	0.85	0.84	0.89	0.85		7
Secondary subfertility		1.79		1.83	2.45		1.74			4
Female factors										
Female age (year)		0.97[a]		1.50[d]	0.9	0.97	0.97			5
Referral status (tertiary care)		0.78								1
Ovulation disorder			0.35							1
Abnormal PCT			0.26				0.23	0.66		3
Pelvic surgery					0.38					1
Tubal defect			0.14	0.5						2
Endometriosis				0.39						1
Ovulation or cervical disorder								0.68		1
Uterine abnormality (UA)								0.45		1
UA and ovulation or cervical disorder								0.3		1
Male factors										
Age, male (year)									0.97	1
Sperm motility (%)	0.91	1.01				0.16[e]	1.01		1.94[g]	5
Degree of motility (good)									0.59	1
Sperm morphology (%)	0.84				1.09	0.78[f]			0.55[h]	4
Sperm concentration (x10^{6})	0.99									1
WHO semen defect				0.47						1
HOS test (%)	0.9									1
Urethritis in history						0.57				1
Fertility problem in man's familiy							0.69			1

LR = logistic regression analysis; CR = Cox proportional hazard regression analysis.

[a] Per year of age ≤31 years; for a female age >31years, an HR of 0.92 has to be calculated for the number of years over 31 years, in addition to the HR for ≤31 years.

[b] Valid if duration of subfertility <24 months.

[c] Valid if duration of subfertility <36 months.

[d] Valid if female age ≤30 years.

[e] Total motility% combined with or without quality of motility (value = 1 when total motility% ≤20 and motility quality <2; 0 otherwise).

[f] Sperm morphology ≤70%.

[g] Valid if sperm motility >85%.

[h] Valid if sperm morphology <40 or ≥90%.

A total of 36 papers were included in our critical appraisal. Some papers discussed an existing model rather than a newly derived model and therefore the number of included models is lower than the number of included papers. There were 12 papers which reported on the prediction of treatment-independent pregnancy. In these papers, nine different prediction models were described. The 4 papers on models for the prediction of pregnancy after IUI reported on 3 different models, and the 20 papers for the prediction of pregnancy after IVF accounted for 17 different models.

The characteristics of the studies in these papers are summarized for the different interventions in Tables I–III. The majority of studies were designed as a prospective cohort study. The inclusion criteria for the patients in the studies on the models for the prediction of treatment-independent pregnancy were generally subfertile couples, evaluated at a secondary or tertiary centre. Anovulation, azoospermia and tubal pathology were the most common exclusion criteria. The participants in the studies for the models of pregnancy after IUI or IVF mostly concerned couples within their first cycle, and in case of IVF, with or without assisted fertilization. A summary of the predictor variables and an estimate of the contribution made by each parameter to the prediction for the different models are shown in Tables IV–VI.

Study quality

An overview of the quality items per intervention is shown in Fig. 4a and b. Patient selection was consecutive in 7 (78%) models for of

Table V Overview of the parameters of the prediction models for pregnancy after IUI (expressed as HRs or ORs)

	Erdem *et al.* (2008)	Steures *et al.* (2004)	Tomlinson *et al.* (1996)	Presence of the parameter in the prediction model (number out of three models)
Type of analysis	LR	LR	LR	
Couple factors				
Duration of subfertility (year)	0.93	0.97	0.98	3
Female factors				
Female age (year)		0.97		1
Tubal defect		0.86		1
Endometriosis		0.71		1
Cervical factor		1.31		1
Unexplained subfertility versus male	0.65			1
Number of follicles	1.79		1.73	2
Endometrial thickness			1.31	1
Cycle number (up to 6)	0.47			1
Male factors				
Sperm motility (%)	1.01		1.05	2

LR = logistic regression analysis; CR = Cox proportional hazard regression analysis.

the treatment-independent pregnancy and in 18 (90%) of the treatment-dependent (IUI and IVF) models. Data collection was prospective in 8 (40%) of the treatment-dependent and in 6 (67%) of the models on treatment-independent pregnancy. Description of the variables for treatment was sufficient in 15 models for pregnancy after treatment (75%) and in 5 (56%) for the treatment-independent models. The description of pregnancy was given in almost comparable numbers of studies. Missing or imputation of missing data was reported for only a few models. Of all models for treatment-independent pregnancy, seven (78%) stated that they used Cox proportional hazards analysis, but only two (22%) described censoring. The amount of interventions between basic fertility work-up and time to pregnancy varied substantially between these models; the basic fertility work-up was clearly described in 67% of studies, and the follow-up duration was adequate in almost all of the studies. Of the treatment-dependent models, diagnosis before treatment was described in 14 (70%), and the protocol of treatment was described in 18 (90%).

Phases of development

The phases of development that the prediction models had passed are shown in Table VII. All models had passed development phase 1, as this was a criterion for inclusion in our review. Of the 29 models for prediction of pregnancy, 6 models had been validated only internally, and only 8 other models had passed the phase of external validation. One model had reached the phase of impact analysis.

Of the eight externally validated models, four models dealt with the prediction of treatment-independent pregnancy (Eimers *et al*., 1994; Collins *et al*., 1995; Snick *et al*., 1997; Hunault *et al*., 2004), one model dealt with the prediction of pregnancy after IUI (Steures *et al*., 2004) and three models dealt with the prediction of pregnancy after IVF (Stolwijk *et al*., 1996; Templeton *et al*., 1996; Hunault *et al*., 2002a). The only model that reached the phase of impact analysis was the model of Hunault *et al*. for the prediction of treatment-independent pregnancy.

Model performance

The performance of the eight models that were externally validated (Eimers *et al*., 1994; Collins *et al*., 1995; Stolwijk *et al*., 1996; Templeton *et al*., 1996; Snick *et al*., 1997; Hunault *et al*., 2002b; Hunault *et al*., 2004; Steures *et al*., 2004) is presented in Table VII. One model for the prediction of treatment-independent pregnancy (Hunault *et al*., 2004) had a poor discrimination (AUC 0.59), but good calibration. The other models for the prediction of treatment-independent pregnancy (Eimers *et al*., 1994; Collins *et al*., 1995; Snick *et al*., 1997) also had a poor discrimination (AUC ranging from 0.59 to 0.67) and did not perform well at calibration.

The one externally validated model for pregnancy after IUI (Steures *et al*., 2004) had poor discrimination (AUC 0.59), but good calibration; it could distinguish between a group with poor chances of pregnancy (0–5%) and a group with good chances of pregnancy (8–11%) (Custers *et al*., 2007). Three models for the prediction of pregnancy after IVF had been externally validated (Stolwijk *et al*., 1996; Templeton *et al*., 1996; Hunault *et al*., 2002a). The model of Templeton *et al*. had a poor discrimination with a c-statistic of 0.63, but differentiated reliably between women with a low and a relatively high probability of success with IVF (Smeenk *et al*., 2000) and was therefore to be considered of good calibration. The model of Stolwijk *et al*. had poor discrimination, with c-statistics ranging from 0.50 to 0.56. Calibration was also poor, because the model could not identify women with a (very) low probability of ongoing pregnancy after IVF (Stolwijk *et al*., 1998). In the most recent validation of the model of Hunault *et al*. for the

Table VI Overview of the parameters of the prediction models for pregnancy after IVF (expressed as HRs or ORs)

	van Weert *et al.* (2008)	Lintsen *et al.* (2007)	Verberg *et al.* (2007)	Carrera *et al.* (2007)	Ottoson *et al.* (2007)	Ferlitsch *et al.* (2004)	Hunault *et al.* (2002)	Bancsi *et al.* (2000)	Stolwijk *et al.* (2000)	Minaretzis *et al.* (1998)	Commenges-Duces *et al.* (1998)	Templeton *et al.* (1996)	Stolwijk *et al.* (1996)[i]	Bouckaert *et al.* (1994)	Haan *et al.* (1991)	Hughes *et al.* (1989)	Nayudu *et al.* (1989)	Presence of the parameter in the prediction model (number out of 17 models)
Type of analysis	LR	CR	LR	LR	LR	LR	LR	LR	CR	LR	LR	LR	LR	LR	LR	LR	LR	
Couple factors																		
Duration of subfertility		0.97													0,64			2
Secondary subfertility	1.4	1.11							1.34[g]									3
Previous succesful IVF											2.12							1
Previous unsuccessful IVF																194[q]		1
Female factors																		
Female age	0.94	[a]		0.89	0.74[d]		0.98	0.95	1.73[h]	0.93	0.28[j]	1.01[k]	0.94	2.05[m]	0.56[p]	1.1[r]		14
Body mass index			0.89		0.88[e]	0.84												3
Unexplained subfertility															1.5			1
Basal FSH					0.55	0.77		0.90										3
Tubal reasons for IVF	0.4											0.93			0.65			3
Tuboperitoneal disease								0.24										1
Endometriosis		1.05[b]																1
Cervical factor subfertility		1.04[b]																1
Previous IVF live birth												2.14						1
Previous IVF preg., no live birth												1.35						1
Previous live birth (no IVF)												1.26						1
Previous preg.(no IVF), no live birth												1.12						1
≥1 previous pregnancy													2.26					1
History of unsuccessful IUI	0.59																	1
Cycle number	1.4																	1
Total amount of rFSH used			0.92[c]															1
Number of ampoules											0.98							1
Antral follicle count				1.15														1
Estradiol stimulation Day 4				1.01														1
hCG																	1.06	1
Pregnancy type follicle																	62[q]	1
Total protein																	10301[q]	1
E_2FD (first day E_2 increase)																	4.5	1
Male factors																		
Sperm motility (mean%)	0.98																	1
Sperm morphology (mean%)	1.01																	1

	van Weert *et al.* (2008)	Lintsen *et al.* (2007)	Verberg *et al.* (2007)	Carrera *et al.* (2007)	Ottoson *et al.* (2007)	Ferlitsch *et al.* (2004)	Hunault *et al.* (2002)	Bancsi *et al.* (2000)	Stolwijk *et al.* (2000)	Minaretzis *et al.* (1998)	Commenges-Duces *et al.* (1998)	Templeton *et al.* (1996)	Stolwijk *et al.* (1996)[l]	Bouckaert *et al.* (1994)	Haan *et al.* (1991)	Hughes *et al.* (1989)	Nayudu *et al.* (1989)	Presence of the parameter in the prediction model (number out of 17 models)
Type of analysis	LR	CR	LR	LR	LR	LR	LR	LR	CR	LR	LR	LR	LR	LR	LR	LR	LR	
Pre-wash total motile count (10^6)	1.0																	1
Antisperm antibodies	3.1																	1
Male factor subfertility (WHO)								0.38							0.49			2
Mild male factor		1.06[b]																1
Severe male factor (ICSI)		1.22[b]																1
Donor sperm											2.4							1
Embryo factors																		
Top-quality embryo availability			2.18															1
Score best embryo					0.6		1.56											2
Score second best embryo					0.78													1
Developmental score							1.30[f]			1.36[i]								2
Morphology score							0.73											1
Nr. of oocytes retrieved			0.93				1.03							3.19[n]				3
Fertilisation ratio at first cycle														3.72[o]				1
Treatment episode															0.86			1

LR = logistic regression analysis; CR = Cox proportional hazard regression analysis.

[a] Hazard ratio (HR): e.g. for 25 years 0.99, for 29 years 1.21, for 35 years 1.0 and for 40 years 0.46.

[b] Tubal pathology was taken as the reference category.

[c] Calculated per units of 75 IU.

[d] Female age in five groups with reference category 1 is 25 to 29 years.

[e] BMI in four groups with reference category group 2 is BMI 18.5 to 25.

[f] Developmental score is further adjusted with a more complex calculation Hunault (2002b)

[g] IVF/ICSI cycles 1–2.

[h] For age ≤30 years; HR is 1.68 for age 31 to 35 years.

[i] OR ranging from 1.36 for a 2-cell good embryo to 2.32 for a 4-cell excellent embryo.

[j] Age ≥38 years.

[k] OR age^2 1006 (beta 0.00501) and OR age^3 1000 (beta 0.00261).

[l] Model A: predictions at the start of the first IVF cycle.

[m] Female age ≥38 years.

[n] <10 oocytes retrieved.

[o] ≥1 oocyte retrieved or more than half of them fertilized.

[p] Female age ≥36 years.

[q] We calculated the ORs of the parameters as OR=exp(β); the βs of the parameters were adopted from the models as stated in the respective papers.

[r] Number of years of the female age over 25 years.

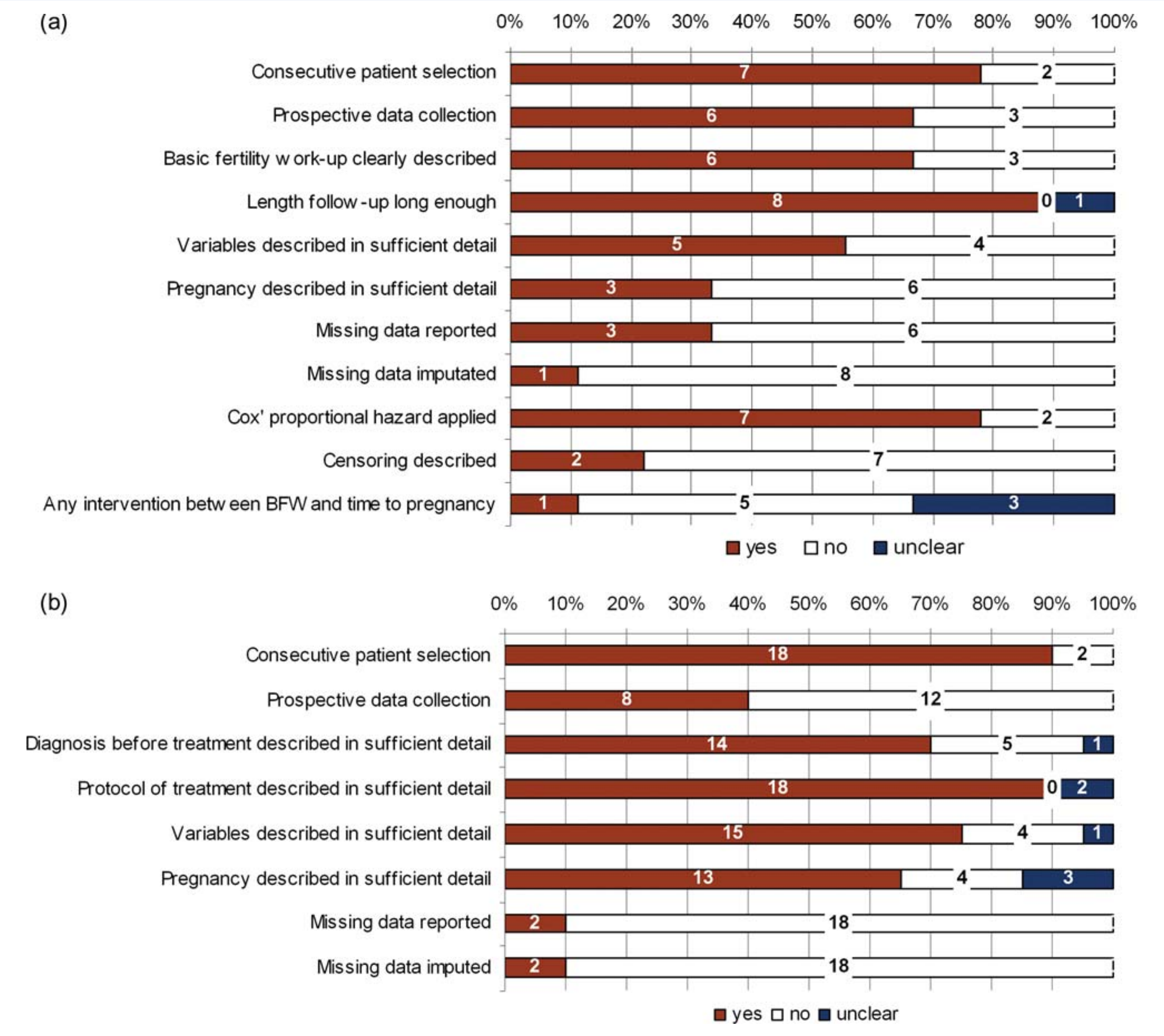


Figure 4 (**a**) Summary of study quality for prediction models for treatment-independent pregnancy. (**b**) Summary of study quality for prediction models for pregnancy after IUI and IVF. The total number of compliant models is presented as the percentage of the stacked bars (100%); numbers in the stack represent the number of models.

prediction of pregnancy after IVF, a c-statistic of 0.63 was reported. However, the reported calibration was poor, because the difference between predicted and observed probabilities was significant ($P < 0.001$) (Hunault *et al*., 2007).

Impact analysis had been performed for the model of Hunault *et al*. for the prediction of treatment-independent pregnancy only, in a large cohort study with an embedded randomized trial (Steures *et al*., 2006). After the basic fertility work-up had been completed, a prognosis for treatment-independent pregnancy was calculated from the model (Hunault *et al*., 2004). The prediction 'model' was transformed into a decision 'rule'. Couples with a good prognosis were counselled for expectant management, whereas couples with a poor prognosis were offered treatment. In the trial, only couples with an intermediate prognosis (a probability of 30–40% for treatment-independent pregnancy within 12 months) were asked to participate in a randomized trial comparing IUI and expectant management. At six months, the ongoing pregnancy rates in both groups were ~25%, which is comparable to the average calculated probability of 30–40% within 12 months.

Clinical application

The populations and outcomes are summarized per intervention in Tables I–III. To illustrate the possible use of the best performing models in clinical practice, we will present a potential clinical application for these models (Supplementary Material, Table S1). A general practitioner has referred a couple, where the 34-year-old

Table VII **Evaluation of model development and model performance**

First author (year)	Phase of development[1]	Specification of model performance at external validation (phase 2b)			
		Report of phase 2b in the manuscript of	Discrimination[2]	Calibration	
				Method of calibration	Result as reported in the paper
Treatment independent					
Jedrzejcak *et al.* (2008)	1	—			
Hunault *et al.* (2004)	2b	van der Steeg *et al.* (2007)	0.59	calibration slope	good
				calibration slope *P*-value	0.82 (95% CI 0.6–1.0) 0.08
	2b	Hunault (2005)	0.59	calibration figure	good
				calibration slope *P*-value	$P = 0.13$[3]
Snick *et al.* (1997)	2b	van der Steeg *et al.* (2007)	—	calibration figure	moderate
				calibration slope *P*-value	0.58 (95% CI 0.4–0.7), $P < 0.01$
	2b	Hunault *et al.* (2004)	0.64–0.65	calibration slope	1.3–1.5
	2b	Snick *et al.* (1997)	0.67	—	
Collins *et al.* (1995)	2b	van der Steeg *et al.* (2007)	—	calibration figure	poor
	2b	Hunault *et al.* (2004)	0.58–0.62	calibration slope	0.6–0.7
	2b	Snick *et al.*(1997)	0.65	—	
Bahamondes *et al.* (1994)	1	—			
Wichmann *et al.* (1994)	1	—			
Eimers *et al.* (1994)	2b	van der Steeg *et al.* (2007)	—	calibration figure	poor
	2b	Hunault *et al.* (2004)	0.59–0.62	calibration slope	0.6–0.8
	2b	Hunault (2002a)	0.62	calibration figure	poor
				calibration slope	0.98 ($P = 0.45$)[4]
Bostofte *et al.* (1993)	1	—			
Bostofte (1987)	1	—			
Intrauterine insemination					
Erdem *et al.* (2008)	1	—			
Steures *et al.* (2004)	2b	Custers *et al.* (2007)	0.59	calibration figure	good
Tomlinson *et al.* (1996)	1	—			
In vitro fertilization					
van Weert *et al.* (2008)	2a	—			
Lintsen *et al.* (2007)	2a	—			
Verberg *et al.* (2007)	2a	—			
Carrera-Rotllan *et al.* (2007)	2a	—			
Ottosen *et al.* (2007)	2a	—			
Ferlitsch *et al.* (2004)	1	—			
Hunault (2002b)	2b	Hunault *et al.* (2007)	0.63	calibration slope *P*-value	$P < 0.001$[4]
		Hunault (2002b)	0.67	Hosmer–Lemeshow	not significant
Bancsi *et al.* (2000)	2a	—			
Stolwijk *et al.* (2000)	1	—			

Continued

Table VII *Continued*

First author (year)	Phase of development[1]	Specification of model performance at external validation (phase 2b)			
		Report of phase 2b in the manuscript of	Discrimination[2]	Calibration	
				Method of calibration	Result as reported in the paper
Minaretzis *et al.* (1998)	1	—			
Commenges-Ducos *et al.* (1998)	1	—			
Templeton *et al.* (1996)	2b	Smeenk *et al.* (2000)	0.63	—	—
Stolwijk *et al.* (1996)	2b	Stolwijk *et al.* (1998)[5]	0.50–0.56[5]	cross-tabulation	poor
Bouckaert *et al.* (1994)	1	—			
Haan *et al.* (1991)	1	—			
Hughes *et al.* (1989)	1	—			
Nayudu *et al.* (1989)	1	—			

[1] The phase of development is defined according to Fig. 1.
[2] Discrimination is reported as the AUC or as the c-statistic.
[3] Results shown for the model without PCT (Hunault *et al.*, 2004).
[4] The model only gave reliable predictions after adjustment of the average live birth rate.
[5] Based on the model I of Stolwijk *et al.* (1996).

woman has primary subfertility of 2 years' duration, to the gynaecologist. The results of the basic fertility work-up revealed no tubal pathology, no uterine abnormalities, but did disclose endometriosis. The post-coital test showed no progressive spermatozoa. The results of the semen analysis showed 40% progressive spermatozoa and no indications for male subfertility. The probability of a treatment-independent pregnancy within 1 year was calculated as 25% using the model developed by Hunault *et al.* (2004). The couple was advised to undergo six treatments of IUI with controlled ovarian stimulation. Using the model developed by Steures *et al.* (2004), one can calculate the probability of pregnancy as 6.3% after one cycle. After unsuccessful IUI treatment, the couple started with IVF. The probability of pregnancy after IVF would be 16%, based on the model of Templeton *et al.* (1996).

Discussion

In this review, of all derived prediction models in reproductive medicine, we identified 29 prediction models. We evaluated the models according to predefined phases of model development and looked systematically at their performance. Only eight models have been externally validated, and only three were found to be of good performance (Templeton *et al.*, 1996; Hunault *et al.*, 2004; Steures *et al.*, 2004). Only the model of Hunault *et al.* for treatment-independent pregnancy had reached the phase of impact analysis.

Our evaluation of prediction models in reproductive medicine was complicated by three major issues. The first issue was the absence of a consensus on which performance measures to use for prediction models and how to interpret them. The AUC of most prediction models, for example, is low but there is a growing recognition that the ROC curve, which plays a central role in evaluating diagnostic models, has limitations in the evaluation of prediction models (Cook, 2007). In contrast to diagnostic accuracy, prognostic accuracy is based on probabilities, and information is lost if the amount of difference between the predicted probabilities and the observed proportion is disregarded. In addition, with some exceptions, such as bilateral tubal obstruction and azoospermia, most couples who attend infertility clinics have some chance of conceiving, whereas on the other hand, even the most fertile couples never have a 100% chance of conception per cycle. Consequently, discrimination will always be imperfect and to use it as a test of a model's performance is not appropriate. Calibration is the most informative way of summarizing the performance of a model (Coppus *et al.*, 2009).

Calibration is evaluated by assessing the level of correspondence between the calculated pregnancy probabilities and the observed proportion of pregnancies. Well-calibrated models are able to classify individuals into clinically useful prognostic strata on the basis of the calculated probabilities of a pregnancy with and without treatment. This is illustrated by the external validation of the Templeton model for the prediction of pregnancy after IVF. The model differentiates between couples with a low and those with a relatively high probability of success after IVF, despite its limited discrimination between couples with and without success, with a c-statistic of 0.63 (Smeenk *et al.*, 2000).

The second issue was the lack of thorough external validation of the prediction models. The majority of the prediction models that were derived for pregnancy after IVF have not yet gone through an external validation. Good performance at external validation is a minimal requirement to be eligible for use in clinical practice. The third issue concerned the generalizability of the models across different patient profiles. The ideal prediction model should guide the gynaecologist to the best policy for a subfertile couple, selecting between expectant management, IUI or IVF. This prediction model should classify couples into groups with different prognoses. Unfortunately, at present it is not possible to calculate these probabilities for an individual couple directly after the completion of the basic fertility work-up. There is not one model for all policies, but there are different models for different policies. These models have been validated in different groups of patients.

The model for the prediction of pregnancy after IVF, for example, was derived and validated in a group of couples at the start of their first IVF cycle. That model has not yet been validated for the prediction of pregnancy after IVF in couples who have just completed the basic fertility work-up, and its performance in that population is unknown.

The models that have been developed in reproductive medicine have reached only the phase of external validation at best, except for the model of Hunault *et al*., which has been used as one of the inclusion criteria in a randomized clinical trial. Further evaluation of model performance after external validation should be encouraged. One of the options is to use the model as a predictive marker in a randomized trial of expectant management versus either IUI or IVF. Such a trial has the advantage that a model can be evaluated for more than one treatment option in the same population, unlike the existing models, which have been evaluated in different patient populations. A second advantage is the fact that one could evaluate the use of the model as a predictive marker in what has been called a marker by treatment interaction design (Sargent *et al*., 2005; Lijmer and Bossuyt, 2009). In such an evaluation, one assesses whether the model is able to accurately identify patients who have better pregnancy chances with one of the treatment options compared with the alternative.

In conclusion, there are now three models with good predictive performance in reproductive medicine (Templeton *et al*., 1996; Hunault *et al*., 2004; Steures *et al*., 2004). These models could be used as a guiding tool in making decisions about fertility treatment in patient couples similar to the development population. Yet, we should encourage further development of these existing models, as well as a more extensive documentation of their contribution to the improvement of the care for individual couples.

Supplementary data

Supplementary data are available at http://humupd.oxfordjournals.org/.

Acknowledgements

The authors thank J. Limpens (PhD, Medical Library, Academic Medical Center) for her assistance in the literature search.

References

Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;**19**:453–473.

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

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

Bostofte E. Prognostic parameters in predicting pregnancy. A twenty-year follow-up study comprising semen analysis in 765 men of infertile couples evaluated by the Cox regression model. *Acta Obstet Gynecol Scand* 1987;**66**:617–624.

Bostofte E, Bagger P, Michael A, Stakemann G. Fertility prognosis for infertile couples. *Fertil Steril* 1993;**59**:102–107.

Bouckaert A, Psalti I, Loumaye E, De CS, Thomas K. The probability of a successful treatment of infertility by in-vitro fertilization. *Hum Reprod* 1994;**9**:448–455.

Carrera-Rotllan J, Estrada-Garcia L, Sarquella-Ventura J. Prediction of pregnancy in IVF cycles on the fourth day of ovarian stimulation. *J Assist Reprod Genet* 2007;**24**:387–394.

Collins JA, Burrows EA, Wilan AR. The prognosis for live birth among untreated infertile couples. *Fertil Steril* 1995;**64**:22–28.

Commenges-Ducos M, Tricaud S, Papaxanthos-Roche A, Dallay D, Horovitz J, Commenges D. Modelling of the probability of success of the stages of in-vitro fertilization and embryo transfer: stimulation, fertilization and implantation. *Hum Reprod* 1998;**13**:78–83.

Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;**115**:928–935.

Coppus SF, van der Veen F, Opmeer BC, Mol BW, Bossuyt PM. Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009; in press.

Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;**45**:562–565.

Custers IM, Steures P, van der Steeg JW, van Dessel TJ, Bernardus RE, Bourdrez P, Koks CA, Riedijk WJ, Burggraaff JM, van der Veen F *et al*. External validation of a prediction model for an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 2007;**88**:425–431.

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. The prediction of the chance to conceive in subfertile couples. *Fertil Steril* 1994;**61**:44–52.

Erdem A, Erdem M, Atmaca S, Korucuoglu U, Karabacak O. Factors affecting live birth rate in intrauterine insemination cycles with recombinant gonadotrophin stimulation. *Reprod Biomed Online* 2008; **17**:199–206.

Ferlitsch K, Sator MO, Gruber DM, Rucklinger E, Gruber CJ, Huber JC. Body mass index, follicle-stimulating hormone and their predictive value in in vitro fertilization. *J Assist Reprod Genet* 2004;**21**:431–436.

Haan G, Bernardus RE, Hollanders JMG, Leerentveld RA, Prak Naaktgeboren FMN. Results of IVF from a prospective multicentre study. *Hum Reprod* 1991;**6**:805–810.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.

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

Hosmer DWLS. *Applied Logistic Regression*, 2nd edn. New York: Wiley and Sons, 2000.

Hughes EG, King C, Wood EC. A prospective study of prognostic factors in in vitro fertilization and embryo transfer. *Fertil Steril* 1989; **51**:838–844.

Hunault CC, Eijkemans MJ, te Velde ER, Collins JA, Habbema JD. Validation of a model predicting spontaneous pregnancy among subfertile untreated couples. *Fertil Steril* 2002a;**78**:500–506.

Hunault CC, Eijkemans MJC, Pieters MHEC, te Velde ER, Habbema JD, Fauser BCJM, Macklon NS. A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. *Fertil Steril* 2002b;**77**:725–732.

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

Hunault CC, Laven JS, van RI, Eijkemans MJ, te Velde ER, Habbema JD. Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum Reprod* 2005; **20**:1636–1641.

Hunault CC, Te Velde ER, Weima SM, Macklon NS, Eijkemans MJC, Klinkert ER, Habbema JD. A case study of the applicability of a prediction model for the selection of patients undergoing in vitro fertilization for single embryo transfer in another center. *Fertil Steril* 2007;**87**:1314–1321.

Jedrzejczak P, Taszarek-Hauke G, Hauke J, Pawelczyk L, Duleba AJ. Prediction of spontaneous conception based on semen parameters. *Int J Androl* 2008;**31**:499–507.

Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009;**62**:364–373.

Lintsen AM, Eijkemans MJ, Hunault CC, Bouwmans CA, Hakkaart L, Habbema JD, Braat DD. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Hum Reprod* 2007; **22**:2455–2462.

McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000;**284**:79–84.

Minaretzis D, Harris D, Alper MM, Mortola JF, Berger MJ, Power D. Multivariate analysis of factors predictive of successful live births in in vitro fertilization (IVF) suggests strategies to improve IVF outcome. *J Assist Reprod Genet* 1998;**15**:365–371.

Nayudu PL, Gook DA, Hepworth G, Lopata A, Johnston WI. Prediction of outcome in human in vitro fertilization based on follicular and stimulation response variables. *Fertil Steril* 1989;**51**:117–125.

Ottosen LD, Kesmodel U, Hindkjaer J, Ingerslev HJ. Pregnancy prediction models and eSET criteria for IVF patients–do we need more information? *J Assist Reprod Genet* 2007;**24**:29–36.

Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;**144**:201–209.

Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005;**23**:2020–2027.

Smeenk JM, Stolwijk AM, Kremer JA, Braat DD. External validation of the Templeton model for predicting success after IVF. *Hum Reprod* 2000; **15**:1065–1068.

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

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

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

Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; **54**:774–781.

Steyerberg EW. *Clinical Prediction Models. A Practical Approach to Development, Validation and Updating*. Springer, 2008.

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

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

Stolwijk AM, Wetzels AM, Braat DD. Cumulative probability of achieving an ongoing pregnancy after in-vitro fertilization and intracytoplasmic sperm injection according to a woman's age, subfertility diagnosis and primary or secondary subfertility. *Hum Reprod* 2000;**15**:203–209.

te Velde ER, Cohlen BJ. The management of infertility. *N Engl J Med* 1999; **340**:224–226.

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

Tomlinson MJ, Missah AJB, Thompson KA, Kasraie JL, Bentick B. Prognostic indicators for intrauterine insemination (IUI): statistical model for IUI success. *Hum Reprod* 1996;**11**:1892–1896.

Tosteson AN, Weinstein MC, Wittenberg J, Begg CB. ROC curve regression analysis: the use of ordinal regression models for diagnostic test assessment. *Environ Health Perspect* 1994;**102**(Suppl. 8):73–78.

van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Bossuyt PM, Hompes PG, van der Veen F, Mol BW. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? *BJOG* 2006;**113**:825–831.

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

van Weert JM, Repping S, van der Steeg JW, Steures P, van der Veen F, Mol BW. A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility. *J Reprod Med* 2008; **53**:250–256.

Verberg MFG, Eijkemans MJC, Macklon NS, Heijnen EMEW, Fauser BCJM, Broekmans FJ. Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF. *Hum Reprod* 2007;**22**:1919–1924.

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

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

World Health Organization. WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. 3rd edn. Cambridge, UK: Cambridge University Press, 1992.

Submitted on December 18, 2008; resubmitted on March 23, 2009; accepted on March 31, 2009