Human Reproduction, Vol.28, No.11 pp. 2972–2980, 2013
Advanced Access publication on August 6, 2013 doi:10.1093/humrep/det315

human reproduction

ORIGINAL ARTICLE *Infertility*

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

L.L. van Loendersloot[1,*], M. van Wely[1], S. Repping[1], P.M.M. Bossuyt[2], and F. van der Veen[1]

[1]Centre for Reproductive Medicine, Department of Obstetrics and Gynaecology, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands [2]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

*Correspondence address. Tel: +31-20-5666199; Fax: +31-20-5669044; E-mail: l.l.vanloendersloot@amc.uva.nl

Submitted on May 22, 2012; resubmitted on June 27, 2013; accepted on July 8, 2013

STUDY QUESTION: Are we able to develop a model to calculate the chances of pregnancy prior to the start of the first IVF cycle as well as after one or more failed cycles?

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

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

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

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

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

LIMITATIONS, REASONS FOR CAUTION: In our study, the outcome of interest was ongoing pregnancy. Live birth may have been a more appropriate outcome, although only 1–2% of all ongoing pregnancies result in late miscarriage or stillbirth. The model was based on data from a single centre.

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

STUDY FUNDING/COMPETING INTEREST(S): This project was funded by the NutsOhra foundation (Grant 1004-179). The NutsOhra foundation had no role in the development of our study, in the collection, analysis and interpretation of data; in writing of the manuscript, and in the decision to submit the manuscript for publication. There were no competing interests.

Key words: assisted reproduction / IVF/ICSI outcome / pregnancy / prediction model

Introduction

Since the introduction of IVF in 1978, over 3.75 million babies have been born worldwide using IVF (ESHRE, 2010). IVF is currently one of the most widely used interventions for infertility. In 2007, 376 971 treatment cycles were reported in 18 European countries, 142 435 cycles in the USA and 56 817 cycles in Australia and New Zealand (CDC, 2007; AIHW, 2007; de Mouzon *et al*., 2012).

IVF is considered as a last resort for all infertile couples regardless of the aetiology of their infertility (NVOG, 1998; NICE, 2004; ESHRE, 2008). In contrast to patients' perceptions, IVF does not guarantee success; almost 50% of couples that start IVF will remain childless, even if they undergo multiple IVF cycles (Moragianni and Penzias, 2010). Given this limited success, it seems logical to offer IVF only to couples with reasonable chances of success and to discontinue treatment when the chances are low and no longer outweigh the burden and costs.

To improve counselling, patient selection and clinical decision-making in IVF, a number of prediction models have been developed in the past (Leushuis *et al.*, 2009). Several models are of limited use as they were developed before current clinical and laboratory protocols were established (Hughes *et al.*, 1989; Nayudu *et al.*, 1989; Haan *et al.*, 1991; Bouckaert *et al.*, 1994; Stolwijk *et al.*, 1996; Templeton *et al.*, 1996; Commenges-Ducos *et al.*, 1998; Minaretzis *et al.*, 1998; Bancsi *et al.*, 2000; Hunault *et al.*, 2002; Carrera-Rotllan *et al.*, 2007). Most models do not include the transfer of frozen–thawed embryos, an essential component of modern-day IVF (Hughes *et al.*, 1989; Nayudu *et al.*, 1989; Haan *et al.*, 1991; Bouckaert *et al.*, 1994; Stolwijk *et al.*, 1996; Templeton *et al.*, 1996; Commenges-Ducos *et al.*, 1998; Minaretzis *et al.*, 1998; Bancsi *et al.*, 2000; Stolwijk *et al.*, 2000; Hunault *et al.*, 2002; Ferlitsch *et al.*, 2004; Carrera-Rotllan *et al.*, 2007; Lintsen *et al.*, 2007; Ottosen *et al.*, 2007; Verberg *et al.*, 2007; van Weert *et al.*, 2008; Nelson and Lawlor, 2011). In Europe alone, almost 86 059 frozen–thawed transfers were performed in 2006 resulting in 10 382 pregnancies—constituting ~15% of all pregnancies achieved in that year (de Mouzon *et al.*, 2010). A number of models calculate pregnancy chances only for the first IVF cycle, while others calculate pregnancy chances after one failed IVF cycle only (Stolwijk *et al.*, 1996; Banerjee *et al.*, 2010). This limits their practical use because the average pregnancy rate is ~29% per cycle, and thus in over 70% of the couples, a decision has to be made whether or not to continue IVF (de Mouzon *et al.*, 2010). We therefore set out to develop a model that would calculate pregnancy chances prior to the start of the first IVF cycle as well as after one or more failed cycles. The model is based on empirical data systematically collected in consecutive IVF patients.

Materials and Methods

Patients

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

All couples in our cohort had been trying to conceive for at least 12 months. They had undergone a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (NVOG *et al.*, 2004). The indication to start IVF or ICSI treatment was determined according to the Dutch IVF guideline (NVOG, 1998). If subfertility was caused by tubal pathology, such as two-sided tubal blockage and severe endometriosis, or severe oligozoospermia (post-wash total motile sperm count <3 million), IVF/ICSI was offered directly (Repping *et al.*, 2002). In the case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia and unexplained subfertility, at least six intrauterine inseminations were performed before IVF/ICSI was offered. In the case of ovulation disorders, mainly caused by polycystic ovary syndrome (PCOS), 12 cycles of ovulation induction were applied before IVF/ICSI was offered.

No medical ethical approval for this research was needed.

Data on clinical diagnoses, IVF protocol and response, and laboratory data on embryo morphology and growth, as well as treatment outcomes for all IVF/ICSI cycles were retrieved from our clinical databases and medical records. Included in the analyses were data on stimulated IVF/ICSI cycles and also from frozen–thawed embryo transfers from these stimulated cycles. We excluded IVF/ICSI cycles that involved oocyte or embryo donation, cycles that used surgically retrieved spermatozoa, cycles from human immunodeficiency virus-positive patients, cycles that involved a modified natural cycle and cycles cancelled due to poor ovarian stimulation, ovarian hyperstimulation syndrome or other unexpected medical or non-medical reasons (Pelinck *et al.*, 2005). Women underwent controlled ovarian hyperstimulation after down-regulation with the GnRH agonist triptorelin (Decapeptyl®) in a long protocol with a midluteal start. Controlled ovarian hyperstimulation was started on cycle Day 5 with recombinant FSH or HMG in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Follicular maturation was induced by 10 000 IU hCG (Pregnyl, Organon). Cumulus–oocyte complexes were recovered by transvaginal ultrasound-guided follicle aspiration 36 h later. Oocytes were inseminated with 10 000 or 15 000 progressively motile spermatozoa (IVF) or injected with a single spermatozoon (ICSI) 2–4 h after follicle aspiration. Embryos were cultured in Human Tubal Fluid (HTF, Cambrex) or G5 medium (Vitrolife) at 37°C and 5% CO_2 in air. Embryo transfer was performed mostly 72 h, and occasionally 96 h, after follicle aspiration with a Wallace catheter (Smiths Medical). Supernumerary embryos of good quality were frozen on Day 4 after follicle aspiration using a slow-freeze protocol. The luteal phase was supported by intravaginal progesterone 200 mg (Utrogestan) two times per day. A hCG blood test was performed 18 days after oocyte retrieval.

Embryos were cultured individually. On each day of development, the number of blastomeres was assessed and each embryo was given a morphological score. For the morphological score, the degree of fragmentation of the embryo and the uniformity of the blastomeres were assessed (Puissant *et al.*, 1987). Based on the degree of fragmentation, embryos were scored as 1 (no fragments), 2 (<20% fragmentation), 3 (20–50% fragmentation) or 4 (>50% fragmentation). If the blastomeres of the embryo were non-uniform in size, the morphological score was reduced by one point, with 4 being the lowest possible score. If on Day 3, the embryo showed signs of compaction, the embryo was scored as a morula and given a grade based on the degree of compaction (score 1: full compaction, score 2: >0–20% compaction and score 3: <20% compaction).

Outcome

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

Data analysis

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

A model was developed to calculate the probability of pregnancy after IVF, including fresh and frozen–thawed embryo transfers from the same cycle. We identified a number of candidate predictors based on a recent systematic review and meta-analysis and on different cohort studies on predictive factors in IVF, reported elsewhere (Templeton *et al.*, 1996; Holte *et al.*, 2007; Lintsen *et al.*, 2007; van Loendersloot *et al.*, 2010). The list of candidate predictors included clinical characteristics available before the start of IVF (female and male age, previous pregnancies, duration of subfertility, indication for IVF and basal FSH), IVF stimulation parameters (initial FSH dose) and laboratory data from the previous failed IVF cycle, if applicable [fertilization method (IVF/ICSI), number of oocytes, number of embryos, embryo quality and number of embryos transferred].

Some of the candidate predictors had missing values. Simple exclusion of couples with missing values for one or more variables commonly causes biased results and decreases statistical efficiency (Greenland and Finkle, 1995). For this reason, we first performed an analysis with missingness indicators and then completed missing values by multiple imputation using the Statistical Package for the Social Sciences (version 18.0) (Greenland and Finkle, 1995). This method uses all available data to impute the missing values based on the correlation between each variable with missing values and all other variables. If candidate predictors had 25% or more missing values, they were excluded from the analyses.

We first checked the linearity of the association between the continuous variables female age, male age, duration of subfertility, basal FSH, initial FSH dose, number of embryos and the logit transformed probability of an ongoing pregnancy using restricted cubic spline functions in univariable logistic regression. We performed similar preliminary analyses for the number of blastomeres, morphological score and embryo implantation. The analysis demonstrated a non-linear association between the continuous variables female age, male age, duration of subfertility, basal FSH, initial FSH dose, number of embryos and ongoing pregnancy after IVF. We therefore transformed all variables to better fit the data. Age was transformed using a polynomial: Age + Age^2 + Age^3. The duration of subfertility was capped at 5 years. Basal FSH was capped at the bottom with values below 10 U/l set at 10 U/l. Initial dose FSH was similarly capped with values above 300 IU coded as 300 IU. The number of embryos was capped at 10 embryos. All embryo morphological scores could adequately be described using linear functions. The number of blastomeres on Day 2 was recoded as the absolute value of the deviation from 4. The number of blastomeres on Day 3 was recoded as the absolute value of the deviation from 8.

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

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

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

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

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

To prevent overfitting and to avoid a too optimistic impression of model performance, a linear shrinkage factor was estimated based on model fit and the number of parameters (Steyerberg, 2009). Coefficients in the model were then corrected by this shrinkage factor.

Performance

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

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

To evaluate any miscalibration, we also fitted a calibration model using logistic regression, with the linear combination of variables in the prediction model as the only variable (Steyerberg *et al.*, 2001; Steyerberg, 2009).

External validation

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

Updating the model

To obtain a model with better precision and stronger validity, we updated the coefficients in the final model after the external validation by re-calibration (Karp *et al.*, 2004; Toll *et al.*, 2008).

Results

We could include data from 1326 couples who had undergone 2621 cycles; of which, 1421 were first IVF cycles, 729 were second IVF cycles, 339 were third IVF cycles and 132 were fourth up to eighth cycles. Two thousand one hundred and ninety-six fresh embryo transfers were conducted 72 h after oocyte retrieval, 202 fresh embryo transfers were conducted 96 h after oocyte retrieval and in 223 cycles, there was no suitable embryo for transfer. There was a total of 903 frozen–thawed cycles, 549 after the first IVF cycle, 229 after the second IVF cycle,

104 after the third IVF cycle and 21 after fourth to eighth IVF cycles. There were 570 ongoing pregnancies from fresh transfers and 82 ongoing pregnancies from frozen–thawed embryo transfer, yielding a total of 652 ongoing pregnancies (24.9% per cycle). The baseline characteristics of the couples are summarized in Table I.

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

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

Table I Baseline characteristics of the cycles included in the development and validation data sets.

	Development set (n = 2621)	Validation set (n = 515)
Clinical characteristics		
Female age years (SD)	35.3 (4.6)	36.71(4.9)
Male age years (SD)	38.4 (6.6)	39.9 (6.7)
Duration of subfertility (years)	3.8 (2.4)	3.9 (2.5)
Type of subfertility		
Primary subfertility (%)	1860 (71%)	346 (67%)
Secondary subfertility (%)	761 (29%)	169 (33%)
Previous ongoing pregnancy	569 (22%)	138 (27%)
FSH (SD)	7.7 (3.6)	7.9 (4.0)
BMI (SD)	24.3 (4.7)	25.1 (5.5)
Indication for IVF		
Unexplained subfertility (%)	518 (20%)	141 (27%)
Tubal pathology (%)	587 (22%)	69 (13%)
Male subfertility (%)	1352 (52%)	227 (44%)
Polycystic ovary syndrome (%)	217 (8%)	18 (3%)
Diminished ovarian reserve (%)	215 (8%)	80 (16%)
Endometriosis (%)	118 (5%)	25 (5%)
Cervical hostility (%)	77 (3%)	10 (2%)
Number of previous failed IVF/ICSI cycles		
0 failed IVF/ICSI cycle (%)	1421 (54%)	312 (61%)
1 failed IVF/ICSI cycle (%)	729 (28%)	121 (23%)
2 failed IVF/ICSI cycles (%)	339 (13%)	62 (12%)
3–7 failed IVF/ICSI cycles (%)	132 (5%)	20 (4%)
IVF stimulation parameters of the previous failed IVF/ICSI cycle		
FSH intial dose (mean, SD)	208 (112)	287 (133)
Type of fertilization		
IVF (%)	1338 (51%)	247 (48%)
ICSI (%)	1239 (47%)	268 (52%)
Embryological data of the previous failed IVF/ICSI cycle		
Number of oocytes (SD)	9.31 (5.6)	9.43 (5.1)
Normal fertilization % (SD)	53.8 (29.3)	45.6 (27.3)
Number of embryos (SD)	5.04 (4.1)	4.54 (3.8)
Mean no. of cells per embryo on Day 3 (SD)	5.33 (2.3)	4.68 (2.2)
No. of 8-cell embryos on Day 3 (SD)	1.15 (1.7)	0.61 (1.0)
No. of morulae on Day 3(SD)	0.09 (0.4)	0.04 (0.3)
No. of embryos with optimal progression (Day 2 4-cell and Day 3 8-cells, SD)	0.90 (1.5)	0.39 (0.9)
Mean morphological score, all embryos Day 3 (SD)	2.2 (0.8)	2.2 (0.9)
Number of frozen embryos (SD)	1.12 (2.4)	0.49 (1.1)
Number of embryos transferred (SD)	1.70 (0.7)	1.86 (1.0)

failed IVF cycles were significantly associated with lower chances of an ongoing pregnancy.

Thirteen predictors were included in the final multivariable logistic regression model. These were the following patient characteristics: female age, duration of subfertility, previous ongoing pregnancy, male subfertility, diminished ovarian reserve, endometriosis, basal FSH and number of failed IVF cycles. We added an interaction term for female age and male subfertility, and one for diminished ovarian reserve and endometriosis. For the calculation of pregnancy chances after the first cycle, we added the following embryo features from the previous cycle (if any) to the patient characteristics: fertilization, number of embryos, mean morphological score per day 3 embryo, presence of 8-cell embryos on Day 3 and presence of morulae on Day 3 (Table II).

There was no significant additional effect of IVF cycle number, nor were there any significant interactions between the identified predictors and cycle number. For this reason, we used the same point estimates for all predictors and included cycle number as a predictor (Supplementary data, Table SI).

The calculated probabilities of an ongoing pregnancy for the 1326 couples in the development set had a wide range: from 0.00 to 0.72, with a mean of 0.25 (Fig. 1). Twenty-five per cent of the cycles had a probability of a pregnancy of <0.17, 25% had a probability between 0.18 and 0.26, 25% a probability between 0.27 and 0.32 and 25% had a probability exceeding 0.33. Four hypothetical cases and the corresponding probabilities are shown as an example in Table III.

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

The validation was performed on data from 440 couples undergoing 515 cycles of IVF. Baseline characteristics are summarized in Table I. The calculated probabilities of an ongoing pregnancy for the 515 cycles

Table II Multivariate analysis for predicting pregnancy chances after an IVF/ICSI cycle.

Predictors	Original model			Updated model	
	β^*	95% CI	*P*-value	β	95% CI
Intercept	26.0950			22.0109	
Patient characteristics					
Age	−2.5792	−4.76 to −0.40	0.01	−2.1909	−4.37 to −0.01
Age^2	0.0851	0.02–0.15	0.01	0.0723	0.01–0.14
Age^3	−0.0009	0.00–0.00	0.00	−0.0008	0.00–0.00
Duration of subfertility[a]	−0.1001	−0.18 to −0.02	0.00	−0.0850	−0.16 to −0.01
Previous ongoing pregnancy	0.2338	0.00–0.47	0.03	0.1986	−0.04 to 0.43
Male subfertility	1.0880	−0.55 to 2.72	0.15	0.9242	−0.71 to 2.56
Diminished ovarian reserve	−0.9239	−1.50 to −0.35	0.00	−0.7848	−1.36 to −0.21
Endometriosis	−0.5635	−1.11 to −0.02	0.03	−0.4786	−1.02 to 0.07
Basal FSH[b]	−0.0798	−0.16 to 0.00	0.04	−0.0678	−0.15 to 0.01
Number of previous failed IVF cycles	−0.2391	−0.43 to −0.05	0.01	−0.2031	−0.40 to −0.01
Age × male subfertility	−0.0322	−0.08 to 0.01	0.14	−0.0274	−0.07 to 0.02
Endometriosis × diminished ovarian reserve	1.7872	0.25–3.32	0.01	1.5181	−0.02 to 3.05
Embryo parameters					
Embryo yes/no after ovum retrieval	0.8503	−0.02 to 1.72	0.04	0.7223	−0.15 to 1.59
Number of embryos after ovum retrieval[c]	0.0610	0.00−0.12	0.02	0.0518	0.00–0.11
Mean morphological score all embryos Day 3	−0.3613	−0.69 to −0.03	0.02	−0.3069	−0.64 to 0.02
8-cell embryo yes/no on Day 3	−0.3315	−0.66 to 0.00	0.03	−0.2816	−0.61 to 0.05
Morulae yes/no on Day 3	0.6219	0.09–1.15	0.01	0.5283	0.00–1.06

CI, confidence interval.
Age^2 = Age squared.
Age^3 = Age to the power of 3.
β^* = corrected beta coefficient for overfit.
[a]Duration of subfertility ≥5 years = 5 years.
[b]Basal FSH ≤10 IE/l = 10 IE/l.
[c]Number of embryos ≥10 = 10 embryos.

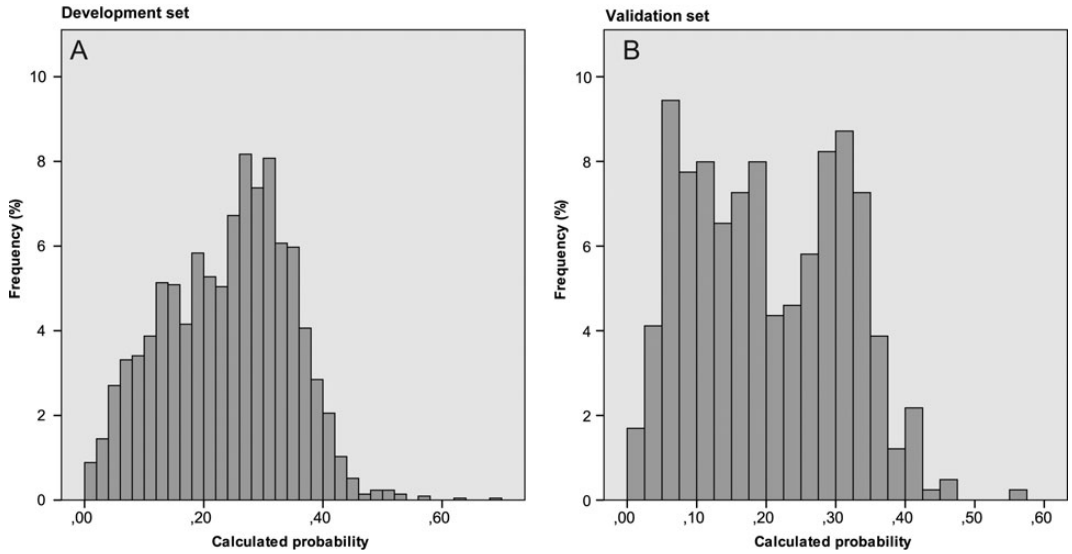

Figure 1 Distribution of the calculated probabilities.

Table III Four hypothetical patients with the calculated probability of an ongoing pregnancy in the subsequent IVF/ICSI cycle.

	Patient A	Patient B	Patient C	Patient D
Age	34	42	36	27
Pregnancy history	None	None	Miscarriage not after IVF	None
Cause of infertility	Unexplained subfertility	Male subfertility and diminished ovarian reserve	Diminished ovarian reserve	Male subfertility
Duration of infertiliy	4 years	4 years	7 years	2 years
Previous IVF cycles	Two	One	None	None
Data from last IVF cycle				
Number of embryos after ovum retrieval	7	4	-	-
Mean morphological score all embryos Day 3	2.0	2.0	-	-
8-cell embryo yes/no on Day 3	yes	no	-	-
Morulae yes/no on Day 3	no	no	-	-
Calculated probability of an ongoing pregnancy	0.25	0.05	0.13	0.37

in the validation set ranged from 0.01 to 0.56, with a mean of 0.22, indicative of a population with more cycles with intermediate and poor prognosis compared with the cycles in the development set (Fig. 1).

The discriminative capacity was similar to that in the development set, with a *c*-statistic of 0.68 (95% CI: 0.63–0.73). The model calibrated well for the first three quintiles, with calculated probabilities in the range from 0.0 to 0.56. The model somewhat underestimated the actual rate in the fourth quintile (calculated probability in the range from 0.26 to 0.32) and overestimated it in the fifth quintile (calculated probability $\geq$0.32). Calibration is summarized in Fig. 2. The slope of the linear predictor (calibration slope) was 0·85 (95% CI: 0.53–1.17), indicating that the calculated probabilities were slightly optimistic: low ones are too low and high ones are somewhat thigh. The calibration intercept was −0.16 (95% CI: −0.59 to 0.28).

The updated final model is summarized in Table II.

Discussion

We developed a prediction model to calculate pregnancy chances during the whole IVF process, both for the first cycle as well as after one or more failed cycles, and taking both fresh embryo transfer and frozen–thawed embryo transfers into account. The model was developed using a careful blockwise building strategy with data systematically collected in consecutive IVF patients. The resulting model produced a range of calculated

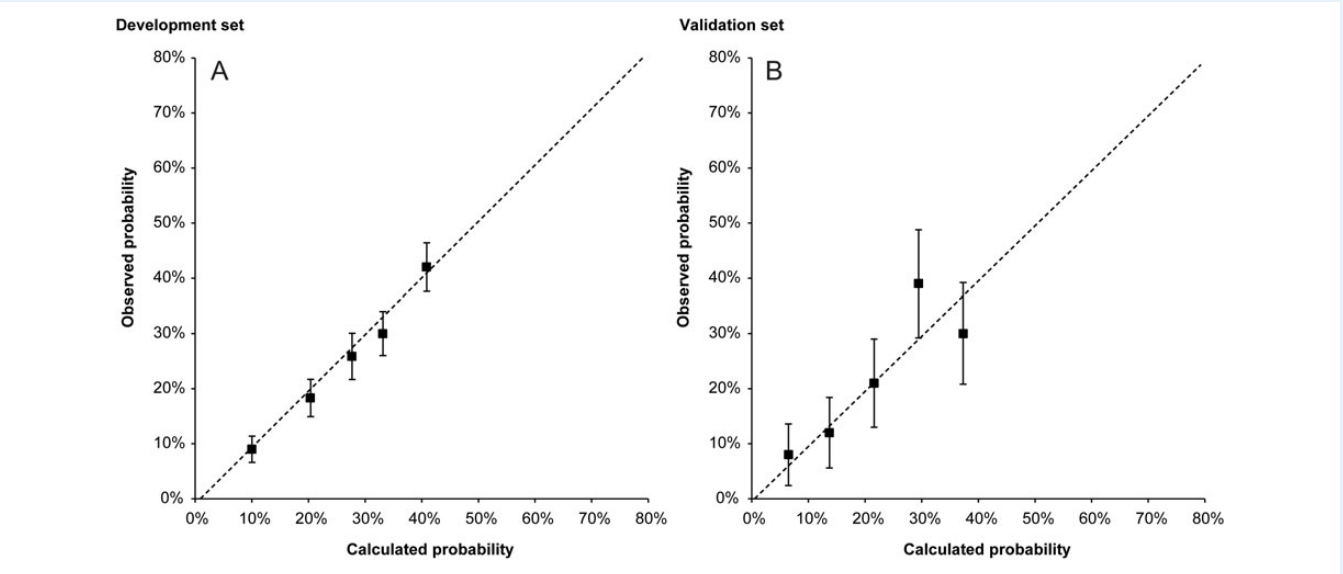

Figure 2 Calibration plots, showing the association between the calculated and observed rates of ongoing pregnancy after IVF/ICSI.

probabilities that were well calibrated, both in the development set and in a separate validation set, which contained data that had not been used for model construction.

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

Live birth as the main outcome for our model would have been ideal. Unfortunately, we did not have these data for all the included cycles. Since only 1–2% of all ongoing pregnancies result in late miscarriage or stillbirth, we do not expect that our model would fundamentally change and we therefore feel that ongoing pregnancy rate is the second most appropriate outcome (Regan and Rai, 2000).

As we used data from a single centre only, the generalizability of the model to other clinics has to be evaluated more extensively in future studies (geographical validation). Centres with higher or lower success rates could use the model after recalibration, by adjusting the intercept to reflect the IVF success rates in their centre (Steyerberg *et al.*, 2004). Such periodic reassessment may also be beneficial within centres, to ensure that calibration is maintained. We have not yet evaluated the impact of the model in counselling individual couples, which is also a topic for additional research.

As is the case for other fertility prediction models, discrimination was less than perfect for our model, expressed by the *c*-statistic (0.68), but the calibration data showed that the model distinguishes well between couples with a poor, moderate and good prognosis in successive IVF cycles. We feel these data on calibration are more relevant for decision-making than discrimination statistics in the assessment of any fertility prediction model. Couples undergoing infertility treatment are not concerned about their chances relative to other couples—which is expressed by discrimination—but worry more about their chances of getting pregnant themselves, which is expressed more adequately by calibration (Steures *et al.*, 2004; Coppus *et al.*, 2009). The calibration was somewhat less optimal in the last two quintiles in the validation set. Couples in the corresponding subgroups have a good prognosis, and can be clearly distinguished from couples with a moderate or poor prognosis. We therefore think that this suboptimal calibration has no real practical relevance as couples with a good prognosis will continue treatment, despite a slightly higher or lower probability.

Since the birth of Louise Brown in 1978, the number of IVF cycles has increased rapidly: in the UK, there were 6650 cycles in 1991 and 57 652 cycles in 2010 (HFEA, 2007; HFEA Human Fertilisation and Embryology Authority, 2010). This increase is not caused by a sudden epidemic of infertility but by increased access to IVF and by expansion of the indications for IVF. At first, IVF was only initiated in couples with bilateral tubal occlusion while later on IVF was also initiated in couples with unexplained subfertility, male subfertility, cervical factor, failed ovulation induction, endometriosis or unilateral tubal pathology (Hull *et al.*, 1985; Hamberger *et al.*, 1998; National Institute for Clinical Excellence, 2004). The major difference between the original indication and the indications for which IVF is conducted nowadays is that the couples with bilateral tubal pathology have a zero chance of natural conception and completely depend on IVF for getting pregnant, whilst couples with the newer indications are subfertile and do have chances of natural conception, which may or may not be better than with IVF. For them, these chances have to be balanced against those with IVF. As IVF can be stressful both physically and emotionally and is not without health risks, subfertile couples should thus be well informed about the chances for success with IVF before each cycle. Unfortunately, at this point, there are no randomized controlled trials comparing IVF with natural conception. Thus, the only way to counsel couples properly is by model-based prognosis.

In the current financial climate, healthcare systems all over the world face dramatic budgets cuts. In the USA alone, these cuts in healthcare cost are expected to amount to 100 billion US dollars, the National Healthcare System in the UK has to reduce its budget by 20 billion pounds and in the Netherlands, these costs reductions are calculated around 5 billion euros (Campbell and Meikle, 2011; Meeus and

Stokmans, 2012; Rappeport, 2012). With these decreases in funding, the IVF budget will inevitably be cut as well. Every fertility specialist should be encouraged to control IVF costs by selecting only those couples for IVF that have a reasonable chance of success that outweighs the burden and health risks of the treatment. At this point, the only way to select couples for IVF is by model-based prognosis. Our model enables an individualized calculation of the chances of ongoing pregnancy with IVF. Based on a couple's specific probability, one can decide whether the chances of an ongoing pregnancy with IVF justify the burden, risks and costs of the treatment.

The use of prediction models in deciding whether couples should receive fertility treatment out of public funding is not new. In The Netherlands and New Zealand, prediction models are used to decide which couples would truly benefit from fertility treatment (i.e. fertility treatment would indeed increases their chance of conception compared with natural conception) and which couples will not benefit (Farquhar *et al.*, 2011).

Before implementing our model in clinical practice, the threshold at which probability to start or to continue treatment should be determined, as this may differ between different stakeholders. To achieve optimal implementation of the model, as shown by a previous implementation study, adequate patient information material should be developed, the organization of regular fertility meetings is necessary, the development of local protocols needs to be further stimulated and the knowledge and communication skills of professionals ought to be improved (van den Boogaard *et al.*, 2012).

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

Supplementary data

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

Authors' roles

L.L.L. contributed to design of the study, acquisition of the data, analysis and interpretation of the data. She also drafted the manuscript. M.W. contributed to design of the study and analysis of the data. She also participated in interpretation of the data and she revised the manuscript critically. S.R. contributed to the design of the study, acquisition of the data, interpretation of the data and to the revisions of the manuscript. P.M.M.B. contributed to the design of the study, analysis and interpretation of the data. He revised the manuscript critically. F.V. contributed to the design of the study, interpretation of the data and to the revisions of the manuscript.

Funding

This project was funded by The NutsOhra foundation (Grant 1004-179). The NutsOhra foundation had no role in the development of our study, in the collection, analysis, and interpretation of data; in writing of the manuscript, and in the decision to submit the manuscript for publication.

Conflict of interest

There were no competing interests.

References

AIHW. Assisted reproductive technology in Australia and New Zealand, 2007. http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442458973 (20 April 2013, date last accessed).

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;**3**:552–557.

Banerjee P, Choi B, Shahine LK, Jun SH, O'Leary K, Lathi RB, Westphal LM, Wong WH, Yao MW. Deep phenotyping to predict live birth outcomes in in vitro fertilization. *Proc Natl Acad Sci USA* 2010;**31**:13570–13575.

Bouckaert A, Psalti I, Loumaye E, de Cooman S, Thomas K. The probability of a successful treatment of infertility by in-vitro fertilization. *Hum Reprod* 1994;**3**:448–455.

Campbell D, Meikle J. £20bn NHS cuts are hitting patients, Guardian investigation reveals. *The Guardian* 2011. http://www.guardian.co.uk/society/2011/oct/17/nhs-cuts-impact-on-patients-revealed (21 April 2012, date last accessed).

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;**9**:387–394.

Centers for Disease Control and Prevention (CDC). *2007 Assisted Reproductive Technology Report*. 2007. http://www.cdc.gov/art/ART2007/PDF/COMPLETE_2007_ART.pdf (21 April 2012, date last accessed).

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;**1**:78–83.

Coppus SF, van der Veen F, Opmeer BC, Mol BW, Bossuyt PM. Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009;**8**:1774–1778.

de Mouzon J, Goossens V, Bhattacharya S, Castilla JA, Ferraretti AP, Korsak V, Kupka M, Nygren KG, Nyboe Andersen A. Assisted reproductive technology in Europe, 2006: results generated from European registers by ESHRE. *Hum Reprod* 2010;**8**:1851–1862.

de Mouzon J, Goossens V, Bhattacharya S, Castilla JA, Ferraretti AP, Korsak V, Kupka M, Nygren KG, Andersen AN. Assisted reproductive technology in Europe, 2007: results generated from European registers by ESHRE. *Hum Reprod* 2012;**4**:954–966.

European Society of Human Reproduction and Embryology (ESHRE). Good clinical treatment in assisted reproduction—an ESHRE position paper. 2008 http://www.eshre.eu/binarydata.aspx?type=doc&sessionId=yyggljojxjtmhjv5555k2peat/Good_Clinical_treatment_in_Assisted_Reproduction_ENGLISH_new.pdf (21 April 2012, date last accessed).

European Society of Human Reproduction and Embryology (ESHRE). *ART Fact Sheet*. 2010. http://www.eshre.eu/ESHRE/English/Guidelines-Legal/ART-fact-sheet/page.aspx/1061 (21 April 2012, date last accessed).

Farquhar CM, van den Boogaard NM, Riddell C, Macdonald A, Chan E, Mol BW. Accessing fertility treatment in New Zealand: a comparison of the clinical priority access criteria with a prediction model for couples with unexplained subfertility. *Hum Reprod* 2011;**11**:3037–3044.

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;**12**:431–436.

Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995; **12**:1255–1264.

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

Hamberger L, Lundin K, Sjogren A, Soderlund B. Indications for intracytoplasmic sperm injection. *Hum Reprod* 1998;**13** (Suppl. 1):128–133.

Human Fertilisation and Embryology Authority (HFEA). A long term analysis of the HFEA Register data 1991–2006. 2007. http://www.hfea.gov.uk/docs/Latest_long_term_data_analysis_report_91–06.pdf (21 April 2012, date last accessed).

Human Fertilisation and Embryology Authority (HFEA). Fertility treatment in 2010: trends and figures. 2010. http://www.hfea.gov.uk/docs/2011–11–16_-_Annual_Register_Figures_Report_final.pdf (21 April 2012, date last accessed).

Holte J, Berglund L, Milton K, Garello C, Gennarelli G, Revelli A, Bergh T. Construction of an evidence-based integrated morphology cleavage embryo score for implantation potential of embryos scored and transferred on day 2 after oocyte retrieval. *Hum Reprod* 2007;**2**:548–557.

Hughes EG, King C, Wood EC. A prospective study of prognostic factors in in vitro fertilization and embryo transfer. *Fertil Steril* 1989;**5**:838–844.

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, Desai KM. Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 1985; **6510**:1693–1697.

Hunault CC, Eijkemans MJ, Pieters MH, te Velde ER, Habbema JD, Fauser BC, Macklon NS. A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. *Fertil Steril* 2002;**4**:725–732.

Karp I, Abrahamowicz M, Bartlett G, Pilote L. Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease. *Am J Epidemiol* 2004;**7**:707–716.

Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol BW, Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**5**:537–552.

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;**9**: 2455–2462.

Meeus T, Stokmans D. Regeerakkoord: miljardenbezuinigingen op zorg, sociale zekerheid en overheid. *NRC* 2012. http://www.nrc.nl/verkiezingen/2012/10/29/regeerakkoord-miljarden bezuinigingen-op-zorg-sociale-zekerheid-en-overheid/ (21 April 2012, date last accessed).

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;**6**:365–371.

Moragianni VA, Penzias AS. Cumulative live-birth rates after assisted reproductive technology. *Curr Opin Obstet Gynecol* 2010;**3**:189–192.

National Institute for Clinical Excellence (NICE) and National Collaborating Centre for Women's, CH. *Fertility: Assessment and Treatment for People with Fertility Problems*. 2004. http://www.rcog.org.uk/files/rcog-corp/uploaded-files/NEBFertilityFull.pdf (21 April 2012, date last accessed).

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;**1**:117–125.

Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. *PLoS Med* 2011;**1**:e1000386

NVOG (Dutch Society of Obstetrics and Gynaecology). Indications for IVF 'Indicaties voor IVF'. NVOG Guideline no. 9, 1998. http://www.nvog-documenten.nl/uploaded/docs/09_indicaties_ivf.pdf (21 April 2013, date last accessed).

NVOG (Dutch Society of Obstetrics and Gynaecology). Guideline-Basic fertility work-up. NVOG Guideline no.1, 2004. http://www.nvog-documenten.nl/index.php?pagina=/richtlijn/item/pagina.php&id=23735&richtlijn_id=518 (21 April 2013, date last accessed).

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;**1**:29–36.

Pelinck MJ, Vogel NE, Hoek A, Arts EG, Simons AH, Heineman MJ. Minimal stimulation IVF with late follicular phase administration of the GnRH antagonist cetrorelix and concomitant substitution with recombinant FSH: a pilot study. *Hum Reprod* 2005;**3**:642–648.

Puissant F, Van Rysselberge M, Barlow P, Deweze J, Leroy F. Embryo scoring as a prognostic tool in IVF treatment. *Hum Reprod* 1987;**8**: 705–708.

Rappeport A. US healthcare groups braced for budget cuts. *Financial Times* 2012. http://www.ft.com/intl/cms/s/0/fbd99bf8-ffbc-11e1–831d-00144feabdc0.html#axzz2RPkN7ysP (21 April 2012, date last accessed).

Regan L, Rai R. Epidemiology and the medical causes of miscarriage. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**5**:839–854.

Repping S, van Weert JM, Mol BW, de Vries JW, van der Veen F. Use of the total motile sperm count to predict total fertilization failure in in vitro fertilization. *Fertil Steril* 2002;**1**:22–28.

Steures P *et al*. Prediction of an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 2004;**1**:45–51.

Steyerberg EW. Clinical Prediction Models. *A Practical Approach to Development, Validation and Updating*. NY, USA: Springer, 2009.

Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001;**1**:45–56.

Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;**16**: 2567–2586.

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; **10**:2298–2303.

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;**1**:203–209.

Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 1996;**9039**:1402–1406.

Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;**11**:1085–1094.

van den Boogaard NM, Musters AM, Bruhl SW, Tankens T, Kremer JA, Mol BW, Hompes PG, Nelen WL, van der Veen F. Tailored expectant management: a nationwide survey to quantify patients' and professionals' barriers and facilitators. *Hum Reprod* 2012;**4**:1050–1057.

van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in *in vitro* fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;**6**:577–589.

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;**4**:250–256.

Verberg MF, Eijkemans MJ, Macklon NS, Heijnen EM, Fauser BC, Broekmans FJ. Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF. *Hum Reprod* 2007;**7**:1919–1924.