Reproductive BioMedicine Online (2014) **29**, 222–230

www.sciencedirect.com
www.rbmonline.com

ARTICLE

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

Laura van Loendersloot [a,*], Madelon van Wely [a], Fulco van der Veen [a], Patrick Bossuyt [b], Sjoerd Repping [a]

[a] *Center for Reproductive Medicine, Women's and Children's Hospital, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands*; [b] *Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands*

* Corresponding author. *E-mail address:* l.l.vanloendersloot@amc.uva.nl (L van Loendersloot).

Laura van Loendersloot graduated from medical school at the University of Amsterdam, The Netherlands. She worked as a fertility doctor and studied for her PhD at the Center of Reproductive Medicine at the Academic Medical Center, University of Amsterdam. She obtained her PhD in 2013 with her thesis entitled 'Predicting IVF outcome'. She is currently a resident in obstetrics and gynaecology at Sint Lucas Andreas Hospital in Amsterdam.

Abstract The selection of embryos based on morphology is still the core of daily laboratory practice in IVF/intracytoplasmic sperm injection. At present, the selection of embryos is primarily based on experience and local protocols. Since an evidence-based ranking strategy for embryos on day 3 is currently lacking, this work constructed a multivariable prediction model to rank embryos according to their implantation potential. A total of 6021 fresh embryo transfers between January 2004 and July 2009 were included, eight potential predictive factors were evaluated and a prediction model was developed using multivariable logistic regression. The model was externally validated with data from couples treated between August 2009 and September 2011 in the same clinic. Five factors were included in the final prediction model: early cleavage, number of blastomeres on days 2 and 3 and morphological score and presence of morula on day 3. With validation, the model showed moderate discriminative capacity (c-statistic 0.70) and calibrated well and was able to distinguish embryos with high ongoing implantation potential from embryos with moderate or low ongoing implantation potential. The model can be used by embryologists as an objective tool to rank embryos according to implantation potential, thereby aiding the selection of embryos for transfer.

KEYWORDS: embryo quality, embryo selection, ICSI, implantation, IVF

http://dx.doi.org/10.1016/j.rbmo.2014.04.016
1472-6483/

Introduction

In the early days of IVF, pregnancy rates per embryo transfer were very low (Edwards and Steptoe, 1983). The only way to increase pregnancy rates at that time was the transfer of large numbers of embryos (Gleicher and Barad, 2006). Over the years, implantation rates per embryo improved and the transfer of these large numbers of embryos led to high multiple pregnancy rates (Human Fertilisation and Embryology Authority (HFEA), 2007; Kingsland et al., 1990; Steptoe et al., 1986). To reduce multiple pregnancy rates, embryo transfer policies restricting the number of embryos to be transferred were introduced. This meant that selection of the embryo(s) with the highest chance of implantation became of paramount importance in IVF, especially since the success rates of cryopreservation of supernumerary embryos were still low at the start of cryopreservation programmes (Human Fertilisation and Embryology Authority (HFEA), 2007).

In the last decade, cryopreservation of embryos has become increasingly successful and this has placed embryo selection in a new context (Human Fertilisation and Embryology Authority (HFEA), 2007; Mastenbroek et al., 2011). Optimal selection of embryos can help to minimize the time to pregnancy by transferring the embryo with the highest implantation potential as early as possible. Transferring only one embryo if there is a high chance of implantation could also help to achieve acceptable pregnancy rates while minimizing the chances of multiple pregnancy.

Ever since the start of IVF, the selection of embryos has been largely based on morphological characteristics of the embryo. Additional methods for embryo selection, such as selection based on chromosomal status (preimplantation genetic screening) and metabolomic profiles of culture media, have been introduced, but upon proper evaluation these methods have been shown to be unable to increase pregnancy rates (Hardarson et al., 2012; Mastenbroek et al., 2007; Vergouw et al., 2011). Morphological selection of embryos thus remains the core of daily laboratory practice in IVF/intracytoplasmic sperm injection (ICSI).

The morphological selection of embryos is largely based on clinical experience and local protocols (ESHRE, 2011; Holte et al., 2007). Several authors have proposed prediction models to rank embryos according to their implantation potential. Unfortunately, most of these models were developed on small data sets and were not externally validated (Giorgetti et al., 1995; Holte et al., 2007; Racowsky et al., 2009, 2011; Steer et al., 1992; Van Royen et al., 1999, 2001; Ziebe et al., 1997).

Based on data from a large cohort of consecutively treated IVF/ICSI patients, this study constructed a new multivariable prediction model to rank embryos according to their implantation potential.

Materials and methods

This study collected data of consecutive IVF/ICSI embryo transfers on day 3 after oocyte retrieval performed between January 2004 and July 2009 in the Centre for Reproductive Medicine of the Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands for the development of the model. For validation of the model, data of embryo transfers performed between August 2009 and September 2011 at the same centre were prospectively collected. All IVF/ICSI embryo transfers were included, regardless of the cycle number of the couple undergoing the treatment.

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

Patients

All couples had been trying to conceive for at least 12 months and underwent a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynecology (NVOG (Dutch Society of Obstetrics and Gynaecology), 2004). The indication to start IVF or ICSI was determined according to the Dutch IVF guideline (NVOG (Dutch Society of Obstetrics and Gynaecology), 1998). If subfertility was caused by tubal pathology, severe endometriosis or severe oligozoospermia (post-wash total motile sperm count <3 million) IVF/ICSI was offered directly (Repping et al., 2002). In case of one-sided tubal pathology, minimal endometriosis, cervical hostility, mild male oligozoospermia or unexplained subfertility, at least six intrauterine inseminations were applied before IVF/ICSI was offered. In case of ovulation disorders, mainly caused by polycystic ovary syndrome, 12 cycles of ovulation induction were applied before IVF/ICSI was offered. The upper female age limit for IVF/ICSI treatment was 43 years.

IVF/ICSI procedures

Women underwent ovarian stimulation after down-regulation with the gonadotrophin-releasing hormone agonist triptorelin (Decapeptyl; Ferring, Hoofdorp, The Netherlands) in a long protocol with a midluteal start. Ovarian stimulation was started on cycle day 5 with recombinant FSH (Gonal-F; Serono Benelux, London, UK; or Puregon; MSD, Oss, the Netherlands) or human menopausal gonadotrophin (Menopur; Ferring) in daily doses ranging from 75 to 450 IU depending on the antral follicle count. Oocyte maturation was induced by 10,000 IU human chorionic gonadotrophin (Pregnyl; MSD). Cumulus–oocyte complexes were recovered by transvaginal ultrasound-guided follicle aspiration 36 h thereafter. 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; Gynotec, Malden, The Netherlands) supplemented with 15% pasteurized plasma protein solution (GPO; Sanquin, Amsterdam, The Netherlands) or G5-PLUS medium (Vitrolife, Göteborg, Sweden) containing human serum albumin at 37°C and 5% (HTF) or 6% (G5-PLUS) CO_2 in air. Embryo transfer was performed on day 3 after oocyte retrieval with a Wallace catheter (Smiths Medical, Rosmalen, The Netherlands). Luteal phase was supported by progesterone intravaginally 200 mg b.i.d. (Utrogestan Besins Healthcare, Brussels, Belgium). An HCG blood test was performed 18 days after oocyte retrieval.

Morphological scoring

Each embryo was cultured individually. Pronuclear scoring was performed 17–22 h after insemination/injection and early

cleavage was scored 23–28 h after insemination/injection. On day 2 (41–46 h after insemination/injection) and day 3 (65–70 h after insemination/injection), the number of blastomeres was assessed and each embryo was given a morphological score, based on the degree of fragmentation of the embryo and the uniformity of the blastomeres (Puissant et al., 1987). The embryo was given a score of 1 (no fragmentation), 2 (<20% fragmentation), 3 (20–50% fragmentation) or 4 (>50% fragmentation). If the blastomeres of the embryo were nonuniform in size for their developmental stage (i.e. the 2-, 4- or 8-cell stage), the morphological score was augmented with one point, with 4 remaining the lowest possible score. If on day 3 the embryo showed signs of compaction, the embryo was scored as a morula and given a grade based on the degree of compaction (1 = full compaction, 2 = 50–99% compaction, 3 = <50% compaction).

Number of embryos transferred

Before July 2006, double-embryo transfer was performed in all women unless there was a medical indication to limit the number of transferred embryos to one. After July 2006, an individualized transfer policy was adopted. Single-embryo transfer was performed in women aged <35 years undergoing their first cycle of IVF/ICSI with at least one top-quality embryo. Double-embryo transfer was performed in women aged <35 who did not have a top-quality embryo in the first cycle, in women aged <35 who failed to get pregnant in their first cycle of IVF/ICSI and in women aged 35–38 years. In women aged ≥39, three embryos were transferred. These strategies were based on a combination of the Practice Committee of the American Society for Reproductive Medicine guidelines and the Belgian embryo transfer legislation (ASRM, 2004, 2006).

Predictors

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

Outcome

The primary outcome was ongoing implantation, defined as the implantation of an embryo that resulted in an ongoing pregnancy with cardiac activity at a gestational age of at least 11 weeks. Ongoing implantation was determined in the case of the presence of one fetus after the transfer of a single embryo, two fetuses if two embryos had been transferred and three fetuses with cardiac activity if three embryos had been transferred.

Statistical analysis

For the development of the model, only embryos with individual traceability were used. These were cycles with single-, double- or triple-embryo transfer that had resulted in either no implantation or transplantation of all transferred embryos. Monozygotic twins were excluded from the analysis. Embryos on which preimplantation genetic screening was performed were also excluded. The embryo was the unit of analysis in model development. This data set is the development data set with traceable embryos.

Some of the candidate predictors had missing values. Simple exclusion of couples with missing values on one or more variables commonly causes biased results and decreases statistical efficiency (Greenland and Finkle, 1995). Missing values in the data were completed by multiple imputation using Statistical Package for Social Sciences version 18.0 (SPSS, USA). 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.

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

For each candidate predictor, this work performed a univariable logistic regression analysis and estimated the corresponding unconditional odds ratios, calculating 95% confidence intervals (CI) and *P*-values. Although hypothesis testing is usually performed with a significance level of 5%, a different significance level for variable section in model building is commonly used, as the incorrect exclusion of a factor would be more deleterious than the inappropriate consideration of one factor too many (Steyerberg et al., 1999). The current work considered all prognostic variables reaching a significance level of 30% in univariable analysis and built a multivariable model with all these variables. For reasons of parsimony, variables were removed from the model if their removal did not significantly reduce model fit, using the generalized likelihood ratio test statistics and a 5% significance level.

This work additionally evaluated whether the choice of ICSI rather than IVF had an effect on embryo selection. This would be the case if there were a significant interaction between the selected treatment and the selected predictive variables.

In deciding between competing expressions of related parameters, Akaike's information criterion (AIC) was used in variable selection (Steyerberg, 2009). The model with the best AIC was selected as the final model. Additionally, all potential predictive factors for interactions were evaluated using an interaction term.

To prevent overfitting and a too optimistic impression of model performance, a linear shrinkage factor was estimated, based on model fit and the number of parameters

(Steyerberg, 2009). Coefficients in the model were then corrected by this shrinkage factor.

Performance of the final model

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

To extrapolate the implantation rates from the data set with individual embryo traceability to the total data set, a correction factor was used. The correction factor was calculated as the ratio of the overall implantation rate (the number of implantations relative to the number of transferred embryos in the total data set) versus the implantation rate in the individual traceability data set (the number of implantations relative to the number of transferred embryos in the individual traceability data set).

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

To evaluate agreement between calculated probabilities of an ongoing implantation and observed proportions of ongoing implantation, the Hosmer–Lemeshow goodness-of-fit test statistic was calculated. In addition, the mean calculated probabilities of ongoing implantation in disjoint subgroups defined by quintiles were compared with the observed ongoing implantation rate in the corresponding groups. The predicted proportion and the observed proportion of ongoing implantations (for traceable embryos) were compared by plotting the observed ongoing implantation rate versus the mean probability in each of the groups, as calculated from the model.

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

External validation of the model

This work performed an external, temporal validation (Steyerberg, 2009). The performance of the model was evaluated for more recent embryo transfers on day 3 after oocyte retrieval, performed between August 2009 and September 2011 in the Centre for Reproductive Medicine of the Academic Medical Centre, the Netherlands. This evaluation was performed in the validation data set with traceable embryos. This work also performed validation on the complete development and validation sets (on a transfer level) containing all transferred embryos. The probability of success in transfer was calculated based on the calculated probabilities of the embryos transferred. The transfers were then ranked in terms of the calculated probabilities and assigned to subgroups, based on quintiles. The predicted proportion and the observed proportion of success (for all transfers) were compared by plotting in each group the observed ongoing implantation rate versus the mean probability, as calculated from the model.

Updating the model

After the external validation, the model was updated based on all available data through recalibration (Karp et al., 2004; Toll et al., 2008). The linear combination of variables in the model were fitted as the only variable in a logistic regression model, using all traceable embryos in the development set and the validation set. Based on the estimated slope and intercept of that model, the intercept and coefficients of the prediction model were adjusted to create a final, updated model.

Results

Between January 2004 and July 2009, 3143 embryo transfers had been performed, transferring a total of 6021 embryos (mean 1.9 embryos per transfer). The 3143 transfers led to at least one viable intrauterine pregnancy of at least 11 weeks in 713 cases (23%). Of the 6021 transferred embryos, 848 implanted: a total ongoing implantation rate of 14%. In 247 transfers (374 embryos), all embryos implanted; in 466 transfers (993 embryos), fewer embryos implanted than were transferred; and in 2430 transfers (4654 embryos), no embryos implanted (Supplementary Figure S1 in the online version at doi:10.1016/j.rbmo.2014.04.016, available online). A total of 5028 transferred embryos had exact traceability and were used further for model development.

Between August 2009 and September 2011, 1666 additional embryo transfers were performed, transferring a total of 3061 embryos (mean 1.8 embryos per transfer). The ongoing pregnancy rate in this validation set was 21% (351/1666). Of the 3061 transferred embryos, 405 implanted: a total ongoing implantation rate of 13%. In 152 transfers (199 embryos), all embryos implanted; in 199 transfers (443 embryos), fewer embryos implanted than transferred; and in 1315 transfers (2419 embryos), no embryos implanted (Supplementary Figure S1 in the online version at doi:10.1016/j.rbmo.2014.04.016). A total of 2618 transferred embryos had exact traceability and were used further for model development.

The baseline characteristics of all embryo transfers and the data sets with traceable embryos are summarized in Supplementary Table S1 in the online version at doi:10.1016/j.rbmo.2014.04.016, both for the development set and for the validation set. One variable had missing values: early cleavage (26% missing).

Analysis with spline functions demonstrated a nonlinear association between the number of blastomeres on days 2 and 3 after oocyte retrieval and ongoing implantation. Both variables were transformed to fit the data better. The number of blastomeres on day 2 was recoded as the absolute value of the deviation from 4; an embryo with six blastomeres was

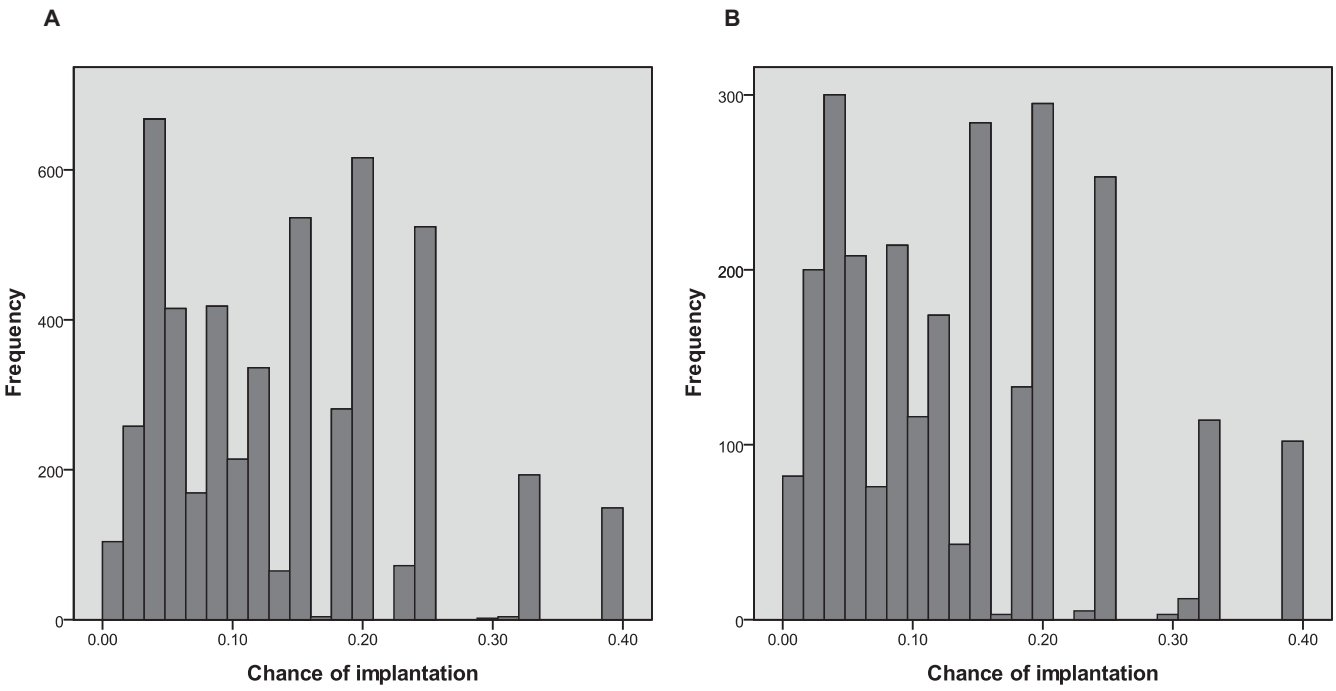

Figure 1 Distribution of the calculated probabilities in the development set and validation set with traceable embryos. (A) Development set. (B) Validation set.

recoded to a score of 2 (6 minus 4) and an embryo with three blastomeres was recoded to a score of 1 (4 minus 3). Similarly, the number of blastomeres on day 3 was recoded as the absolute value of the deviation from 8. All embryo morphology scores could adequately be described by linear functions (Supplementary Figure S2 in the online version at doi:10.1016/j.rbmo.2014.04.016).

In univariable analysis, early cleavage, number of blastomeres on days 2 and 3, morphological score on days 2 and 3 and progression from 4 blastomeres in day-2–8 blastomeres on day 3 were found to be significantly associated (all $P < 0.01$ at a 30% significance level) with ongoing implantation (Supplementary Table S2 in the online version at doi:10.1016/j.rbmo.2014.04.016). The pronuclear score and the presence of morula on day 3 were not significantly associated with ongoing implantation.

After the removal of variables from the resulting multivariable model without loss in goodness of fit, four factors were found to be significantly associated with ongoing implantation: number of blastomeres on day 2, number of blastomeres on day 3, the morphological score on day 3 and presence of morula on day 3 (all $P < 0.001$). Early cleavage was not significantly associated with ongoing implantation. These factors were included in the final multivariable logistic regression model. None of the evaluation interactions between these terms was statistically significant and no interaction terms were included in the final model. There were no significant interactions between treatment (ICSI or IVF) and the predictive variables. This means that the ranking of embryos is not affected by the use of ICSI rather than IVF. The goodness-of-fit of the final model was compared with a model that consisted of day-3 parameters only (i.e. number of blastomeres and morphological score on day 3). The final model fitted the data significantly better than the other models (likelihood ratio test; $P < 0.001$).

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

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

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

The performance of the model was also evaluated for all embryo transfers. The model calibrated well both in the complete development set and validation set (Figure 2C and D).

The updated final model and a simplified embryo score are presented in Table 1. The total score can be calculated with the following formula: Total score = 103 + (2 × early cleavage (yes = 1, no = 0)) + (–3 × blastomeres on day 2 deviating from 4) + (–3 × blastomeres on day 3 deviating from 8 (morula = 0)) + (–5 × morphological score on day 3 (morula = 0)) + (–11 × morula on day 3 (yes = 1, no = 0)).

The higher the total score, the higher the ongoing implantation potential of the embryo. Table 2 depicts the

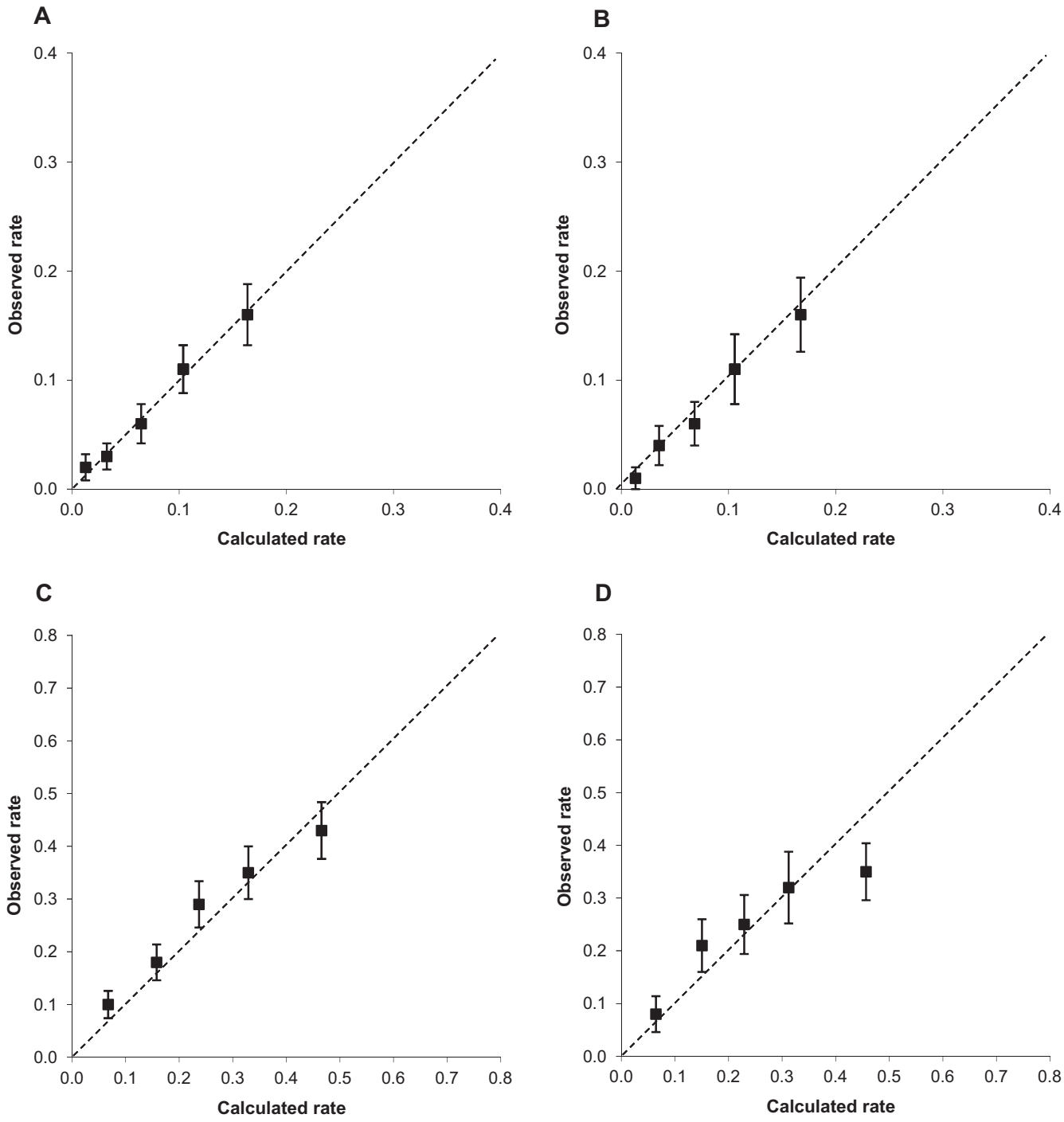

Figure 2 Calibration plots showing associations between calculated and observed ongoing embryo implantation rates. (A) Development set with traceable embryos. (B) Validation set with traceable embryos. (C) Validation of complete development set. (D) Validation of complete validation set. To calculate the actual implantation rate (IR) for panels A and B, the calculated probabilities for the development and validation sets need to be corrected with a factor 1.9 (IR in total data set / IR in development set = 1.9). (A) Corrected calculated probabilities of an ongoing implantation ranged from 0.00 to 0.39 (mean 0.14). (B) Corrected calculated probabilities of an ongoing implantation ranged from 0.00 to 0.39 (mean 0.14). (C and D) These are the corrected calculated probabilities.

hypothetical case of a couple that has 10 embryos after an IVF/ICSI cycle. Their embryos are ranked according to their implantation potential, as calculated with the model. An embryo with early cleavage, 4 blastomeres on day 2 and 8 blastomeres on day 3 with a morphological score of 1 on day 3 has a total score of 100 ($103 + (2 \times 1) + (-3 \times 0) + (-3 \times 0) + (-5 \times 1) + (-11 \times 0)$) (Table 2, embryo 10). The embryo with the highest total score has the highest chance of implantation compared with the other nine embryos.

Table 1 Multivariable analysis.

Predictor	Updated model		Embryo score
	Beta (β)	P-value	
Intercept	−1.0579		103
Early cleavage	0.2492	NS	2
No. of blastomeres on day 2 (deviation from 4)[a]	−0.3324	<0.001	−3
No. of blastomeres on day 3 (deviation from 8)[b]	−0.3128	<0.001	−3
Morphological score on day 3[c]	−0.5305	<0.001	−5
Morula on day 3[d]	−1.1940	<0.001	−11

NS = not significant ($P > 0.05$).
[a]No. of blastomeres = absolute value (no. of blastomeres − 4).
[b]No. of blastomeres = absolute value (no. of blastomeres − 8); morula = 0.
[c]Morula = 0.
[d]Presence of morula = 1; no morula = 0.

Table 2 Hypothetical example of embryo ranking 3 days after oocyte retrieval.

Embryo	Early cleavage	No. of blastomeres on day 2[a]	No. of blastomeres on day 3[b]	Morphological score on day 3	Morula on day 3	Total score
1	No	2	3	3	No	67
2	No	3	5	3	No	76
3	No	4	12	2	No	81
4	No	5	7	2	No	87
5	No	3	NA	NA	Yes	89
6	Yes	4	9	2	No	92
7	No	4	8	2	No	93
8	Yes	4	NA	NA	Yes	94
9	No	4	8	1	No	98
10	Yes	4	8	1	No	100

NA = not applicable.
[a]No. of blastomeres transformed to absolute value (no. of blastomeres − 4).
[b]No. of blastomeres transformed to absolute value (no. of blastomeres − 8); morula = 0.

Discussion

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

One of the strengths of this study is that it evaluated seven embryo predictors in consecutive transfers, using only embryos with exact traceability. The model was developed in a large data set (>6000 embryos) and validated thoroughly using more recent data, collected at the same clinic after the development of the model.

The moderate discriminative capacity implies that the model is not able to distinguish perfectly between embryos with small differences in ongoing implantation potential. Yet perfect prediction is not the goal of this embryo selection model: the primary goal is not to predict with absolute certainty whether an embryo will implant, but to rank embryos based on their ongoing implantation potential within a single treatment cycle of a couple. Although the ideal outcome of the model would be the number of live births, ongoing implantation rate was used as the outcome of interest. Whether implementation of this model ultimately improves ongoing implantation rates and time to pregnancy has not yet been evaluated and is a topic for future research. Since <2% of all ongoing pregnancies result in late miscarriage or still birth, the model is not expected to change fundamentally when using the number of live births as outcome measure (Regan and Rai, 2000).

When scoring the embryos, there was a range in timing of scoring of maximally 5 h. This range could potentially lead to different classification of embryos. Yet despite the time difference of 5 h, early cleavage is still a significant predictive factor in the model. If the time difference had been less and data would more homogeneous, early cleavage would probably have been an even better predictor.

Also the model had an acceptable discriminative capacity and calibrated perfectly even after external validation in spite of this time difference. This work used data of a single centre only, so the generalizability of the model to other clinics has to be evaluated more extensively in future studies (geographical validation). As the aim of the model was to rank the embryos acquired after an IVF/ICSI cycle of a couple and not to calculate the exact implantation rate of an individual embryo, higher or lower implantation rates should not influence the performance of the model. Over the years that the data were collected, there were two significant changes in the study centre: the embryo transfer policy shifted more towards single-embryo transfer (as seen by the lower number

of embryos transferred in the validation set) and there was a switch in culture media (HTF to Vitrolife). Despite these changes, the model still had near-perfect calibration and acceptable discriminative capacity both after internal and external validation, indicating that these changes did not affect the performance of the model.

As indicated in the introduction, several other embryo implantation models have been developed in the past. Previous studies used much smaller data sets and not all used data of embryos with exact traceability of the individual embryos (Giorgetti et al., 1995; Holte et al., 2007; Racowsky et al., 2009; Steer et al., 1992; Van Royen et al., 1999, 2001; Ziebe et al., 1997). Several studies did not validate their model (Giorgetti et al., 1995; Holte et al., 2007; Racowsky et al., 2009; Steer et al., 1992; Ziebe et al., 1997). As prediction models may not perform as well in a new data set as in the development set, external validation of models is essential to support general applicability of the model (Steyerberg, 2009). It also enables further fine tuning of the model by updating the weight of each variable.

An additional problem with some other embryo implantation models is that they included patient characteristics into the model (Racowsky et al., 2009, 2011). As patient characteristics for each of these embryos are identical (all embryos are from the same couple), it is misleading to include these characteristics. They will seem to improve model fit, without actually contributing to the ranking potential. Some of these patient characteristics, such as age, are much stronger predictors than embryo parameters, so including these in a model would result in an overestimation of the discriminative capacity of the model in distinguishing between embryos of the same woman in which female age is identical.

There is also an ongoing discussion as to whether it is better to score embryos only on day 3 or on both day 2 and day 3 (Racowsky et al., 2009). The current work compared the final model to a model that consisted of day-3 parameters only (i.e. number of blastomeres and morphological score). The final model fitted the data significantly better than the day-3 model ($P < 0.001$). Therefore, an implantation model that includes day-2 and day-3 variables has better predictive performance than models that exclude these variables. Whether scoring embryos on day 2 is cost-effective and results in higher pregnancy rates has to be evaluated in future studies.

The association between the five identified embryo factors and embryo implantation is biologically plausible: embryos that demonstrate early cleavage are known to be more likely to implant because they are likely to cleave more evenly, which is strongly correlated with a lower incidence of mitotic chromosomal errors and therefore a higher chance of implantation (Hardarson et al., 2001). The current analyses showed that faster- and slower-cleaving embryos have a lower chance of implantation. The biological explanation could be that embryos that cleave directly into more than 2 cells and embryos that cleave too slowly or too fast also have significantly more chromosomal abnormalities (Hardarson et al., 2006; Magli et al., 2007). In addition, mouse embryos that cleave faster have recently been shown to have greater perturbations in genomic imprinting and metabolic marker expression (Market Velker et al., 2012). The degree of fragmentation of an embryo is strongly correlated with chromosomal mosaicism and embryos that display fragmentation are less likely to implant (Munne and Cohen, 1998).

Early cleavage (day 1) and the number of blastomeres on day 2 are important predictors in this model, implying that embryo selection should not be solely based on embryo parameters assessed on day 3. Culturing embryos individually and scoring them on each day therefore allows for better embryo selection. Newly developed real-time embryo monitoring systems enable the continuous monitoring of embryos and could assist in accurate determination of the timing of all cleavages (Kirkegaard et al., 2012). In the absence of sufficiently large randomized clinical studies, it remains to be elucidated whether embryo selection using dynamic parameters improves clinical outcome or whether it has additional predictive capacity for implantation. Before such randomized controlled studies are performed, morphological selection based on daily evaluation of the embryos seems to remain at the core of current laboratory practice in IVF/ICSI. In addition, it is important to mention that trials that compare real-time embryo analysis to standard daily monitoring also indirectly compare the quality of different incubators, because incubators used for real-time analysis differ from incubators used for standard IVF/ICSI. Thus, to truly assess the (cost–)benefit of real-time analysis, a trial should be performed comparing real-time to standard analysis using the same incubator for both arms of the trial.

Combining a multivariable model that takes into account both the prognostic profile of the patient and the ranking of the embryos could be an important step towards a patient-tailored embryo transfer strategy. Such a combined model would enable calculation of ongoing implantation chances and multiple ongoing implantation chances. This is especially relevant in the situation where a decision has to be made to transfer one or two embryos. Currently, embryo quality is mostly dichotomized into top-quality and non-top-quality embryos. This study shows that embryos can be ranked more precisely based on their ongoing implantation potential and that dichotomizing embryo quality is most likely an oversimplification of reality.

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

Acknowledgements

The authors thank S Mastenbroek for helpful discussions.

Appendix: Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.rbmo.2014.04.016.

References

ASRM, 2004. Guidelines on the number of embryos transferred. Fertil. Steril. 82, 773–774.

ASRM, 2006. Guidelines on number of embryos transferred. Fertil. Steril. 86, S51–S52.

Edwards, R.G., Steptoe, P.C., 1983. Current status of in-vitro fertilisation and implantation of human embryos. Lancet 2, 1265–1269.
ESHRE, 2011. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. Hum. Reprod. 26, 1270–1283.
Giorgetti, C., Terriou, P., Auquier, P., Hans, E., Spach, J.L., Salzmann, J., Roulier, R., 1995. Embryo score to predict implantation after in-vitro fertilization: based on 957 single embryo transfers. Hum. Reprod. 10, 2427–2431.
Gleicher, N., Barad, D., 2006. The relative myth of elective single embryo transfer. Hum. Reprod. 21, 1337–1344.
Greenland, S., Finkle, W.D., 1995. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am. J. Epidemiol. 142, 1255–1264.
Hardarson, T., Hanson, C., Sjogren, A., Lundin, K., 2001. Human embryos with unevenly sized blastomeres have lower pregnancy and implantation rates: indications for aneuploidy and multinucleation. Hum. Reprod. 16, 313–318.
Hardarson, T., Selleskog, U., Reismer, E., Wigander, A., Wennerstrom, S., Westin, C., Ahlstrom, A., Hanson, C., 2006. Zygotes cleaving directly into more than two cells after 25–27 h in culture are predominantly chromosomally abnnormal. Hum. Reprod. 21, i102.
Hardarson, T., Ahlstrom, A., Rogberg, L., Botros, L., Hillensjo, T., Westlander, G., Sakkas, D., Wikland, M., 2012. Non-invasive metabolomic profiling of Day 2 and 5 embryo culture medium: a prospective randomized trial. Hum. Reprod. 27, 89–96.
Holte, J., Berglund, L., Milton, K., Garello, C., Gennarelli, G., Revelli, A., Bergh, T., 2007. 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. 22, 548–557.
Human Fertilisation and Embryology Authority (HFEA). A long term analysis of the HFEA Register data 1991–2006. 2007. Available at: External link <http://www.hfea.gov.uk/docs/Latest_long_term_data_analysis_report_91-06.pdf.pdf> (accessed 02.02.14).
Karp, I., Abrahamowicz, M., Bartlett, G., Pilote, L., 2004. Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease. Am. J. Epidemiol. 160, 707–716.
Kingsland, C.R., Steer, C.V., Pampiglione, J.S., Mason, B.A., Edwards, R.G., Campbell, S., 1990. Outcome of triplet pregnancies resulting from IVF at Bourn Hallam 1984–1987. Eur. J. Obstet. Gynecol. Reprod. Biol. 34, 197–203.
Kirkegaard, K., Agerholm, I.E., Ingerslev, H.J., 2012. Time-lapse monitoring as a tool for clinical embryo assessment. Hum. Reprod. 27, 1277–1285.
Magli, M.C., Gianaroli, L., Ferraretti, A.P., Lappi, M., Ruberti, A., Farfalli, V., 2007. Embryo morphology and development are dependent on the chromosomal complement. Fertil. Steril. 87, 534–541.
Market Velker, B.A., Denomme, M.M., Mann, M.R., 2012. Loss of Genomic Imprinting in Mouse Embryos with Fast Rates of Preimplantation Development in Culture. Biol. Reprod. 86, 1–16.
Mastenbroek, S., Twisk, M., van Echten-Arends, J., Sikkema-Raddatz, B., Korevaar, J.C., Verhoeve, H.R., Vogel, N.E., Arts, E.G., de Vries, J.W., Bossuyt, P.M., Buys, C.H., Heineman, M.J., Repping, S., van der Veen, F., 2007. In vitro fertilization with preimplantation genetic screening. N. Engl. J. Med. 357, 9–17.
Mastenbroek, S., van der Veen, F., Aflatoonian, A., Shapiro, B., Bossuyt, P., Repping, S., 2011. Embryo selection in IVF. Hum. Reprod. 26, 964–966.
Munne, S., Cohen, J., 1998. Chromosome abnormalities in human embryos. Hum. Reprod. Update 4, 842–855.
NVOG (Dutch Society of Obstetrics and Gynaecology). Indications for IVF 'Indicaties voor IVF'. NVOG Guideline no. 9, 1998. Available at: External link <http://www.nvog-documenten.nl/uploaded/docs/09_indicaties_ivf.pdf> (accessed 02.02.14).
NVOG (Dutch Society of Obstetrics and Gynaecology). Guideline-Basic fertility work-up. NVOG Guideline no. 1, 2004. Available at: External link <http://nvog-documenten.nl/index.php?pagina=/richtlijn/pagina.php&fSelectTG_62=75&fSelectedSub=62&fSelectedParent=75> (accessed 02.02.14).
Puissant, F., Van Rysselberge, M., Barlow, P., Deweze, J., Leroy, F., 1987. Embryo scoring as a prognostic tool in IVF treatment. Hum. Reprod. 2, 705–708.
Racowsky, C., Ohno-Machado, L., Kim, J., Biggers, J.D., 2009. Is there an advantage in scoring early embryos on more than one day? Hum. Reprod. 24, 2104–2113.
Racowsky, C., Stern, C., Gibbons, W., Behr, B., Pomeroy, K., Biggers, J.D., 2011. National collection of embryo morphology data into Society for Assisted Reproductive Technology Clinic Outcomes Reporting System: associations among day 3 cell number, fragmentation and blastomere asymmetry, and live birth rate. Fertil. Steril. 95, 1985–1989.
Regan, L., Rai, R., 2000. Epidemiology and the medical causes of miscarriage. Baillieres Best Pract. Res. Clin. Obstet. Gynaecol. 14, 839–854.
Repping, S., van Weert, J.M., Mol, B.W., de Vries, J.W., van der Veen, F., 2002. Use of the total motile sperm count to predict total fertilization failure in in vitro fertilization. Fertil. Steril. 78, 22–28.
Steer, C.V., Mills, C.L., Tan, S.L., Campbell, S., Edwards, R.G., 1992. The cumulative embryo score: a predictive embryo scoring technique to select the optimal number of embryos to transfer in an in-vitro fertilization and embryo transfer programme. Hum. Reprod. 7, 117–119.
Steptoe, P.C., Edwards, R.G., Walters, D.E., 1986. Observations on 767 clinical pregnancies and 500 births after human in-vitro fertilization. Hum. Reprod. 1, 89–94.
Steyerberg, E.W., 2009. Clinical Prediction Models. A Practical Approach to Development, Validation and Updating, Springer, New York.
Steyerberg, E.W., Eijkemans, M.J., Habbema, J.D., 1999. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis 3. J. Clin. Epidemiol. 52, 935–942.
Steyerberg, E.W., Eijkemans, M.J., Harrell, F.E., Jr., Habbema, J.D., 2001. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med. Decis. Making 21, 45–56.
Toll, D.B., Janssen, K.J., Vergouwe, Y., Moons, K.G., 2008. Validation, updating and impact of clinical prediction rules: a review. J. Clin. Epidemiol. 61, 1085–1094.
Van Royen, E., Mangelschots, K., De Neubourg, D., Valkenburg, M., Van de Meerssche, M., Ryckaert, G., Eestermans, W., Gerris, J., 1999. Characterization of a top quality embryo, a step towards single-embryo transfer. Hum. Reprod. 14, 2345–2349.
Van Royen, E., Mangelschots, K., De Neubourg, D., Laureys, I., Ryckaert, G., Gerris, J., 2001. Calculating the implantation potential of day 3 embryos in women younger than 38 years of age: a new model. Hum. Reprod. 16, 326–332.
Vergouw, C.G., Kieslinger, D.C., Kostelijk, E.H., Hompes, P.G., Schats, R., Lambalk, C.B., 2011. Metabolomic profiling of culture media by near infrared spectroscopy as an adjunct to morphology for selection of a single day 3 embryo to transfer in Ivf: a double blind randomised trial. Fertil. Steril. 96, S3.
Ziebe, S., Petersen, K., Lindenberg, S., Andersen, A.G., Gabrielsen, A., Andersen, A.N., 1997. Embryo morphology or cleavage stage: how to select the best embryos for transfer after in-vitro fertilization. Hum. Reprod. 12, 1545–1549.

Declaration: The authors report no financial or commercial conflicts of interest.

Received 14 August 2013; refereed 12 April 2014; accepted 14 April 2014.