Human Reproduction, Vol.28, No.5 pp. 1210–1220, 2013
Advanced Access publication on March 5, 2013 doi:10.1093/humrep/det030

human reproduction

ORIGINAL ARTICLE *Embryology*

Objective way to support embryo transfer: a probabilistic decision

L. Gianaroli[1,2,*], M.C. Magli[1,2], L. Gambardella[3], A. Giusti[3], C. Grugnetti[2], and G. Corani[3]

[1]SISMER, Reproductive Medicine Unit, via Mazzini 12, Bologna 40138, Italy [2]IIRM International Institute for Reproductive Medicine, via Sant'Anna 1, Sorengo 6924, Switzerland [3]DalleMolle Institute for Artificial Intelligence (IDSIA) and SUPSI, Manno 6928, Switzerland

*Correspondence address. Tel: +30-051-307307; E-mail: luca.gianaroli@sismer.it

Submitted on August 16, 2012; resubmitted on November 22, 2012; accepted on January 18, 2013

STUDY QUESTION: Is it feasible to identify factors that significantly affect the clinical outcome of IVF-ICSI cycles and use them to reliably design a predictor of implantation?

SUMMARY ANSWER: The Bayesian network (BN) identified top-history embryos, female age and the insemination technique as the most relevant factors for predicting the occurrence of pregnancy (AUC, area under curve, of 0.72). In addition, it could discriminate between no implantation and single or twin implantations in a prognostic model that can be used prospectively.

WHAT IS KNOWN ALREADY: The key requirement for achieving a single live birth in an IVF-ICSI cycle is the capacity to estimate embryo viability in relation to maternal receptivity. Nevertheless, the lack of a strong predictor imposes several restrictions on this strategy.

STUDY DESIGN, SIZE, DURATION: Medical histories, laboratory data and clinical outcomes of all fresh transfer cycles performed at the International Institute for Reproductive Medicine of Lugano, Switzerland, in the period 2006–2008 ($n = 388$ cycles), were retrospectively evaluated and analyzed.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Patients were unselected for age, sperm parameters or other infertility criteria. Before being admitted to treatment, uterine anomalies were excluded by diagnostic hysteroscopy.

To evaluate the factors possibly related to embryo viability and maternal receptivity, the class variable was categorized as pregnancy versus no pregnancy and the features included: female age, number of previous cycles, insemination technique, sperm of proven fertility, the number of transferred top-history embryos, the number of transferred top-quality embryos, the number of follicles >14 mm and the level of estradiol on the day of HCG administration. To assess the classifier, the indicators of performance were computed by cross-validation. Two statistical models were used: the decision tree and the BN.

MAIN RESULTS AND THE ROLE OF CHOICE: The decision tree identified the number of transferred top-history embryos, female age and the insemination technique as the features discriminating between pregnancy and no pregnancy. The model achieved an accuracy of 81.5% that was significantly higher in comparison with the trivial classifier, but the increase was so modest that the model was clinically useless for predictions of pregnancy. The BN could more reliably predict the occurrence of pregnancy with an AUC of 0.72, and confirmed the importance of top-history embryos, female age and insemination technique in determining implantation. In addition, it could discriminate between no implantation, single implantation and twin implantation with the AUC of 0.72, 0.64 and 0.83, respectively.

LIMITATIONS, REASONS FOR CAUTION: The relatively small sample of the study did not permit the inclusion of more features that could also have a role in determining the clinical outcome. The design of this study was retrospective to identify the relevant features; a prospective study is now needed to verify the validity of the model.

WIDER IMPLICATIONS OF THE FINDINGS: The resulting predictive model can discriminate with reasonable reliability between pregnancy and no pregnancy, and can also predict the occurrence of a single pregnancy or multiple pregnancy. This could represent an effective support for deciding how many embryos and which embryos to transfer for each couple. Due to its flexibility, the number of variables in the predictor can easily be increased to include other features that may affect implantation.

STUDY FUNDING/COMPETING INTERESTS: This study was supported by a grant, CTI Medtech Project Number: 9707.1 PFLS-L, Swiss Confederation. No competing interests are declared.

Key words: embryo implantation / embryo morphology / probability of pregnancy / single embryo transfer

Introduction

Embryo implantation represents a critical step in the reproductive process during which the blastocyst penetrates the endometrium to establish an interface between the growing fetus and the maternal circulation (Guzeloglu-Kayisli *et al*., 2009). For a successful implantation to occur a viable embryo must establish a synchronized dialog with a receptive endometrium (Simon *et al*., 2000).

Owing to the complexity of this process, several factors can determine implantation failure including chromosomal abnormalities, which are especially frequent in female gametes in a manner that is closely related to age (Gianaroli *et al*., 2005a; Kuliev *et al*., 2005; Munné *et al*., 2007; Gianaroli *et al*., 2010), a poor uterine receptivity, sometimes due to asynchrony between the embryonic development and the endometrium (Norwitz *et al*., 2001; Margalioth *et al*., 2006; Swain and Smith, 2011), and factors associated with ART interventions comprising the potential adverse effects of ovarian stimulation (Santos *et al*., 2010) and of *in vitro* culture systems (Jones *et al*., 2001; Horsthemke and Ludwig, 2005; Dumoulin *et al*., 2010) on embryo development.

Looking at clinical data where single implantation frequently results from the transfer of two or more embryos of comparable morphology, it is clear that a large proportion of failed implantations must be ascribed to the embryo. Despite numerous advances in culture systems and the introduction of more refined tools for the classification of the embryo's developmental capacity (Katz-Jaffe *et al*., 2006; Brison *et al*., 2007; Jones *et al*., 2008a, b; Wong *et al*., 2010; Meseguer *et al*., 2011), the lack of a strong predictor of embryo viability imposes severe restrictions to the general implementation of elective single-embryo transfer. In comparative trials between elective single- and elective double-embryo transfers in selected patient groups, the resulting live birth rates are comparable only when the transfer of cryopreserved embryos in subsequent cycles is included (Pandian *et al*., 2009; Gelbaya *et al*., 2010).

These considerations lead to the conclusion that the key requirement for achieving a single live birth within an IVF-ICSI cycle is the capacity to estimate the embryo's potential for further development in relation to maternal receptivity, which is a general term including factors associated with the patient characteristics and the treatment cycle. Morphological criteria have been, and still are, routinely used to select embryos for transfer or cryopreservation, and comprise the evaluation of pronucleate oocytes, the occurrence of early cleavage at 25–27 h post-insemination, scoring systems for day 2–3 embryos (including number of cells and synchrony of cleavage as well as the degree and pattern of fragmentation) and grading systems for morulas (progression of compaction) and blastocysts (including the timing of blastocyst formation and the degree of expansion as well as the characteristics of the inner cell mass and trophectoderm cell lineages) (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011).

These criteria brought the definition of top-quality embryos. They are currently the most predictive markers of embryo developmental competence and are widely used in most IVF laboratories (Kovacic *et al*., 2004; Volpes *et al*., 2004; Gianaroli *et al*., 2007; Magli *et al*., 2007; Racowsky *et al*., 2011). Other studies have shown that the predictive capacity can be further improved by combining the scores obtained by the morphology of the same embryo at different stages (De Placido *et al*., 2002; Lan *et al*., 2003; Sjöblom *et al*., 2006; Loi *et al*., 2008; Brezinova *et al*., 2009) as well as by using the information coming from morphokinetic observations (Wong *et al*., 2010; Meseguer *et al*., 2011). Nevertheless, reliable prediction of which embryos will implant remains an open problem, as prognostic models generally achieve only limited accuracy (Saith *et al*., 1998; Beuchat *et al*., 2008).

Based on the assumption that for a successful implantation to occur a viable embryo must meet a receptive uterus, this study proposes a predictor of pregnancy that was designed by analyzing a series of cycles with known clinical outcomes. This was done to identify the factors that, being specifically related to embryo viability and maternal receptivity, were shown to affect the clinical outcome. The resulting predictor is able to discriminate with reasonable reliability not only between pregnancy and no pregnancy, but also between the occurrence of a single pregnancy or multiple pregnancy. This could represent an effective support to decide for each couple how many embryos and which embryos to transfer.

Materials and Methods

Patients

A total of 388 cycles performed at the International Institute for Reproductive Medicine of Lugano, Switzerland, between 2006 and 2008 were included in the study. Patients were unselected for age, sperm parameters or other infertility criteria. The average maternal age was 36.3 $\pm$ 4 years, and the average number of previous IVF-ICSI cycles was 2.2 $\pm$ 1.7. Before being admitted to treatment, uterine anomalies were excluded by diagnostic hysteroscopy (Gianaroli *et al*., 2005b).

Ovarian stimulation was achieved by long down-regulation ovarian simulation protocols (Ferraretti *et al*., 2004). Oocytes were retrieved transvaginally via ultrasound guidance at 34–36 h after HCG administration, and cultured in HTF (SAGE Cooper Surgical, Inc., Pasadena, CA) supplemented with 5% human serum albumin (HSA, SAGE) at 37°C in a 5.3% CO_2 humidified gas atmosphere. Insemination was performed ~5 h later with conventional IVF in 143 cycles (37%) or ICSI in 245 cycles (63%) depending on sperm indices and the couple's reproductive history.

At 16 h after insemination, oocytes were checked for the presence of pronuclei and polar bodies. Regularly fertilized oocytes were cultured individually and scored daily at 40, 64, 88 and 112 h after insemination. According to the National regulation on IVF, a maximum of three fertilized oocytes per cycle were left in culture. Embryo transfer was performed between Days 2 and 5 depending on the clinical workload: 77 cycles on Days 2, 141 on Days 3, 95 on Day 4 and 75 on Day 5.

A clinical pregnancy was defined as the presence of a gestational sac with fetal heartbeat at 7–8 weeks. All pregnancies were followed-up to term. The implantation rate was calculated by the ratio between the number of gestational sacs with fetal heart beat and the number of transferred embryos.

The study was supported by a grant, CTI Medtech, Project Number: 9707.1 PFLS-L, Swiss Confederation.

Pronucleate oocytes and embryo grading

Pronucleate oocyte morphology was assessed by observing (i) the shape and position of pronuclei, (ii) the distribution and size of nucleoli within pronuclei and (iii) the position of polar bodies with respect to the longitudinal axis of pronuclei. According to previous studies, the configurations having juxtaposed and centrally located pronuclei with a symmetrical dis-

tribution of large size nucleoli and polar bodies next to each other are considered as top quality having the highest chances of normal chromosome status, embryo development and implantation (Gianaroli et al., 2003; Gianaroli et al., 2007).

Day 2–3 embryos were assessed by recording the number and morphology of nuclei and blastomeres, the percentage of fragmentation in the perivitelline space and the presence of compaction (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). Day 4 embryos were classified based on the number of cells and grade of compaction and day 5 embryos were assessed according to the grade of expansion associated with inner cell mass and trophectoderm morphology (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011).

The definition of top-quality embryos was as follows:

(i) Day 2: embryos with four regular cells and fragmentation <10%;
(ii) Day 3: embryos with eight regular cells and fragmentation <10%;
(iii) Day 4: embryos at the morula stage with fully compacted blastomeres;
(iv) Day 5: embryos at the blastocyst stage with inner cell mass and trophectoderm.

Embryos graded as top quality throughout all stages were graded as top history. In addition, day 4–5 embryos were graded as top history also if they had been graded as top quality in all stages but one, provided that the single non-top assessment did not occur on the day of transfer (see later for details).

Statistical analysis

Statistical classifiers are mathematical models, which predict the value of a categorical target variable (*the class*) on the basis of the values of other variables (*the features*). The class variable was categorized as pregnancy versus no pregnancy; the features included the following variables that normally represent the main characteristics of a treatment cycle: the woman's age, previous IVF-ICSI cycles, insemination technique, quality of sperm, number of transferred top-history embryos, number of transferred top-quality embryos, number of follicles >14 mm and level of estradiol on the day of HCG administration. For each cycle, the classifier estimated the probability of pregnancy and no pregnancy. The most probable outcome was regarded as a prediction and the *accuracy* of the classifier, namely the proportion of successful predictions, was measured. Since in IVF-ICSI cycles, the occurrence of no pregnancy is two to four times more frequent than pregnancy, a trivial classifier, which always predicts no pregnancy, can achieve high accuracy without conveying any useful information. Therefore, the accuracy of the developed classifiers was checked against that of the trivial classifier.

Given the shortcomings of accuracy, which is very sensitive on the class imbalance and allows the trivial classifier to obtain a good evaluation, the classifiers were also evaluated through the AUC, namely the area under the receiver operating characteristic curve. For example, in the case of two women, of whom only one is destined to become pregnant, the AUC corresponds to the probability that the model will assign a higher probability of pregnancy to that woman compared with the second. The performance of the model is graded as poor if the AUC lies between 0.5 and 0.7, fair if the AUC lies between 0.7 and 0.8, good if the AUC exceeds 0.8, while an AUC of 1 implies perfect discrimination (Leushuis et al., 2009).

As final evaluation, the *calibration* of the predictions was checked, namely the correspondence between the computed and the observed proportion of pregnancies.

To assess the classifier, the indicators of performance were computed by *cross-validation* (Witten et al., 2011). The original sample was split into k parts; the model was derived on $(k-1)$ parts of the data (*training data*) and the predictions were evaluated on the remaining part (*test data*). The procedure was repeated several times until each part was used as the test set. The k parts were stratified, namely they contained a similar proportion of pregnancies and no pregnancies.

Decision trees

As a first step, data were analyzed through a decision tree that was expected to provide good accuracy, and to represent at the same time a model easy to be interpreted by IVF-ICSI professionals.

The decision tree tried to discover the patterns, which link the features with the occurrence of pregnancy or no pregnancy. The model expressed the discovered patterns between class and features as a set of understandable rules. Since *feature selection* was performed as pre-processing step (Witten et al., 2011), not all the features that were initially postulated were included, with the least predictive ones being removed. Feature selection was aimed at making the model better performing and more interpretable.

A Bayesian network model for decision support

Bayesian networks (BNs) are probabilistic models suited for data analysis, representation of expert knowledge and probabilistic modelling (Koller and Friedman, 2009). The BN of this study was developed with the aim of: (i) discriminating with more success than the decision tree not only between no pregnancy and pregnancy, but also between no pregnancy, single pregnancy, twin pregnancy or higher order pregnancy and (ii) serving as a support for the decision of which embryos to transfer, by assessing how the probabilities of having i ($i = 0, 1, 2, 3$) gestational sacs varies with the number and the type of the transferred embryos.

The BN model assumed a viable embryo and a receptive uterus to be necessary for a successful implantation (Zhou and Weinberg, 1998; Simon et al., 2000; Roberts, 2007). In detail, if the uterus is not receptive, there will be no pregnancy; if the uterus is receptive and there is one viable embryo, a single pregnancy will follow; if there are two viable embryos and a receptive uterus, a twin pregnancy will follow, and so on. Moreover, the BN model assumed that each embryo implants independently of any others.

As shown in Fig. 1, the BN model was represented as a graph, where each node characterizes a variable. Contrary to the decision tree, whose discrimination threshold is learned from data, the BN model requires the variables to be discrete; therefore variables were divided into the categories shown in Table I. An arc connecting two nodes represents a relationship between the two variables; in Fig. 1, the maternal receptivity (intended as the predisposition to pregnancy and including the uterine receptivity) depends on the age of the woman and on the number of her previous IVF-ICSI cycles; the viability of the embryo depends on its score (non-top, top-quality, top history) and on the insemination technique; the pregnancy depends on both the maternal receptivity and viability of each embryo. The structure of the model was selected, among possible alternative layouts, adopting the Bayesian Information Criterion.

The BN model required estimating probability distributions for each node, namely the marginal probabilities for all the nodes with no predecessors (e.g. previous cycles, age, score of the embryos) and the conditional probabilities for all other nodes, given all possible combinations of their direct predecessors. For instance, it was necessary to estimate with which probability the embryo would be viable given all the possible combinations of *technique* (IVF or ICSI) and *score* (no-transfer, non-top, top-quality, top history); with which probability the woman would be receptive, given all the possible combinations of *age* (<34, 34–40, >40 years) and previous IVF-ICSI cycles (≤2, >2). The age intervals were

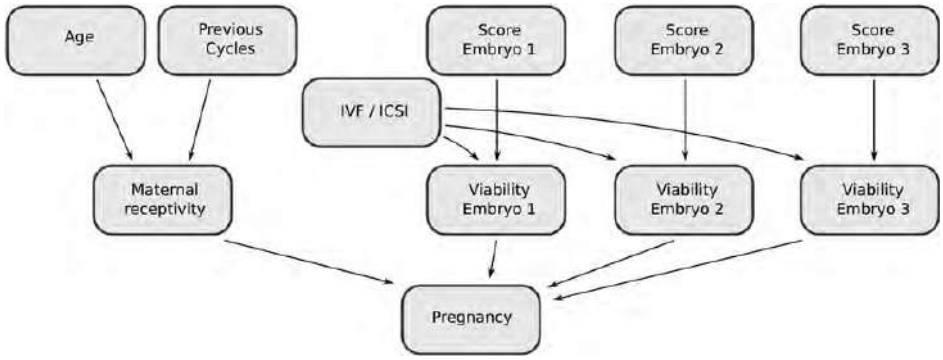

Figure 1 The BN model. The uterine receptivity (intended as predisposition to pregnancy) depends on the number of previous IVF-ICSI cycles and on age. The viability of each embryo depends on both its grade (non-top, top-quality, top history) and the insemination technique. Pregnancy depends on both maternal receptivity and viability of each embryo.

Table I Division of variables for the BN

Variable	Categories
Age (years)	<34, 34–40, >40
Previous IVF-ICSI cycles	0–2, >2
Grade of each embryo	No transfer; non-top; top; top history
Insemination technique	IVF, ICSI
Maternal/uterine receptivity	Receptive; non-receptive
Embryo viability	Viable; non-viable
Pregnancy	No; single; twin; triple

decided according to the indications derived from the retrospective analysis of more than 10 000 internal cycles in relation to maternal receptivity.

The main difficulty in making these estimations was that both maternal receptivity and embryo viability were only *partially observed*. In particular, the uterus viability was observed (as positive) only for the cases with pregnancy; for the cases with no pregnancy, it could not be ascertained whether the uterus was not receptive, the embryos were not viable, or both. Similarly, the viability of the embryos was observed (as positive) only for the cases when all the transferred embryos implanted. To address this problem, the Expectation–Maximization algorithm was used, which is suited to estimate the parameters of a model from a data set with partially observed variables (Zhou and Weinberg, 1998; Roberts, 2007; Roberts *et al.*, 2010).

Once all the conditional probabilities were estimated, the model was ready for predictions; also in this case cross-validation was implemented and the cases used for testing the model predictivity were different from those used for estimating the model parameters. While issuing the prediction for a new case, the BN model was provided with information about embryo grades, woman age, insemination technique and number of previous IVF cycles, while embryo viabilities and uterine receptivity were unknown to the model, as in real clinical practice. On the basis of the available information, the BN model inferred the probability of each embryo being viable and of the uterus being receptive, and consequently estimated the probability of single and multiple pregnancies.

The network of Fig. 1 is handled between one and three embryos to be transferred, as this is the limit imposed by the national regulation on IVF.

Taking decision with expected utility

Based on the variables that were found to be related to pregnancy from the analysis of the 388 cycles, this study provided some examples of the decision, which would have been suggested by applying the criterion of *expected utility* on top of the probabilities estimated by the BN model. The utility is a score, which represents the desirability of the different outcomes. In reproductive medicine, the most desirable clinical outcome is a singleton pregnancy; therefore, the corresponding utility should be set at the top level having a score of 1, while no pregnancy should be set at the bottom level with a score of 0. The occurrence of multiple pregnancy could be set at any intermediate value, including 0, or even at a negative value like in this study, according to the doctor's experience. As an illustrative example, let us assume the utility of no pregnancy, single pregnancy and multiple pregnancy to be, respectively, 0, 1 and 0.5. Let us also assume the probability of no pregnancy, single pregnancy and multiple pregnancy for a certain category of patient to be, respectively, 0.2, 0.5 and 0.3. The expected utility of embryo transfer is obtained by weighing the utility of the different outcomes by their probability ($0 \times 0.2 + 1 \times 0.5 + 0.5 \times 0.3 = 0.65$). However, the chances of pregnancy also vary in relation to the embryos selected for the transfer and on their number and morphology. The probabilities computed by the model can be exploited to select the set of embryos whose transfer maximizes the expected utility.

As a first step, it was required to associate a *utility score* with the possible outcomes of an IVF cycle: no pregnancy, single pregnancy, twin pregnancy or triple pregnancy. The score of no pregnancy (U_0) and single pregnancy (U_1) were set, respectively, *as* $U_0 = 0$ and $U_1 = 1$. The utilities U_2 and U_3 of twin and triple pregnancy, respectively, should reflect the wishes of the patient and/or the recommendation of the doctor. To demonstrate the applicability of the model, we hypothesized the following utilities in relation to the woman's body build that was defined by the body mass index (BMI) according to the WHO obesity classification (World Health Organization, 2000):

- Overweight (BMI >25 kg/m^2): $U_2 = 0.5$, $U_3 = -1$
- Normal (BMI 18.5–25 kg/m^2): $U_2 = 0.2$, $U_3 = -2$
- Underweight (BMI <18.5 kg/m^2): $U_2 = -1$, $U_3 = -2$.

These utility scores were provided as an example; they could be set at levels different from those set here, including leaving a single implantation as the only positive utility. Depending on the clinical experience, other parameters which have an effect on the utility related to the chances of

implantation could have been chosen instead of BMI or in addition to it. The corresponding figures can be adjusted for any single case.

The expected utility of transferring a certain cohort of embryos was $O_i = 3P_iU_i$ where P_i was the probability of having *i* gestational sacs computed by the BN model. In this way, the model could serve as a decision support, suggesting the transfer of the set of embryos resulting in the highest expected utility.

Results

Clinical outcomes

The overall clinical pregnancy rate was 19.8% (77/388), including both single ($n = 60$) and twin pregnancies ($n = 17$), while no triple pregnancies were generated. The pregnancy rate was higher for conventional IVF (26.7%) than for ICSI (16.3%, $P < 0.01$), and was negatively correlated with the number of previous cycles, being, respectively, 21 and 16.5% for women having experienced 0–2 or more than two cycles, although the difference in these pregnancy rates was not significant at a 95% level. A negative correlation was also found with age, going from 28% (25/90) for women aged <34 years to 20% (48/241) for women between 34 and 40 years and 7% (4/57) for women older than 40 years.

The most discriminative variable for predicting a clinical pregnancy was the number of transferred top-history embryos. The pregnancy rate varied from 14% (47/332) to 53% (30/56; $P < 0.01$) depending on whether the cohort contained 0–1 or 2–3 top-history embryos. The 95% confidence intervals of the pregnancy rate for transfers with two to three top-history embryos was 41–65% with even the lower bound of this confidence interval being about twice the average pregnancy rate in the whole data set.

If embryos were categorized only on the basis of their last grade irrespective of their history, the *top-quality* label would be assigned to all the top-history embryos and to those, which were top in the last observation without being top history. In this case, the pregnancy rate calculated in the data set would be of, respectively, 11 or 42% for transfers with 0–1 or 2–3 top embryos. Therefore, the top-history score was more discriminative than the usual top-quality score.

No correlation was found between the number of top-history embryos and female age. The percentage of top-history embryos decreased with the day of culture, going from 35% on Day 2 to 10% on Day 5.

Predictions with the decision tree

For the decision tree, the following features were identified as fundamental to discriminate between pregnancy and no pregnancy: (i) the number of transferred top-history embryos, (ii) the female age and (iii) the fertilization technique (IVF or ICSI). The remaining features were discarded probably because of a certain degree of redundancy (e.g. number of transferred top-quality embryos versus number of top-history embryos; semen of proven fertility versus IVF/ICSI) or because of limited predictive power (estradiol, number of follicles >14).

As shown in Fig. 2, the decision tree checked first the number of transferred top-history embryos; if this was 0 or 1, no-pregnancy was predicted. If the number of transferred top-history embryos was 2 or 3, pregnancy was predicted in the case of conventional IVF, while for ICSI pregnancy was predicted in women younger than 37 years. The number of top-history embryos was recognized as the most discriminative feature, being the first to be tested within the tree. Therefore, the tree predicted pregnancy for cycles characterized by two or three top-history embryos and conventional IVF, or by

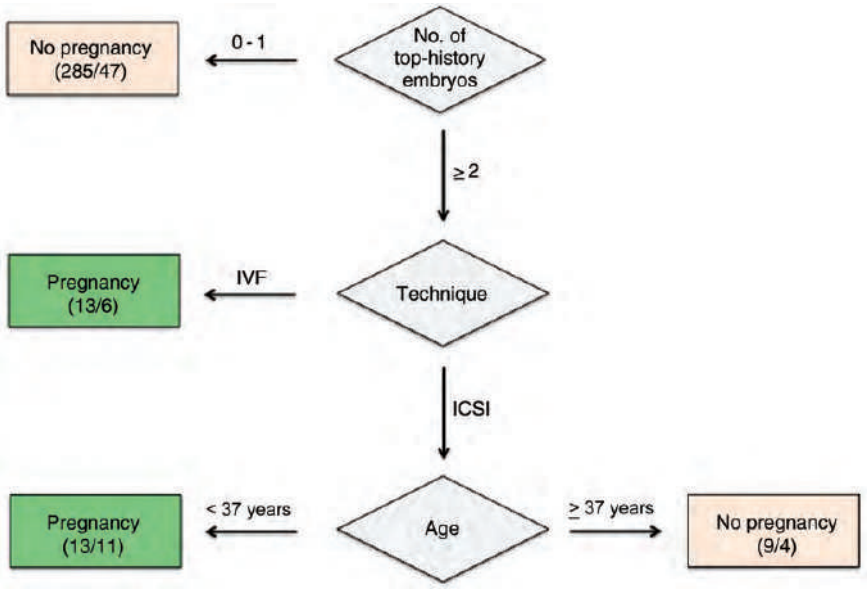

Figure 2 The decision tree with the features enclosed into rhomboidal nodes. The rectangular nodes indicate a point where the prediction is given, with the numbers representing the cases in which the prediction was correct and incorrect, respectively.

two or three top-history embryos, ICSI and woman age <37 years. Overall, 43 cycles matched this criteria, and the pregnancy rate for these cycles was 60% (26/43), roughly three times the average. This confirmed the ability of the decision trees to identify the conditions favouring pregnancy as well as the importance of top-history embryos in determining implantation.

The decision tree achieved an accuracy of 81.5%, as measured by cross-validation. This was significantly higher than the 80.2% achieved by the trivial classifier ($P < 0.05$), but the difference was quite narrow in practical terms. The AUC of the decision tree was 0.65, for which the discrimination ability of the model was quite modest. The performance was even worse when the discrimination between no pregnancy, single pregnancy and twin pregnancy was attempted after re-deriving the model and computing the AUC three times, once for each class (no pregnancy, single pregnancy and twin pregnancy); the AUCs were, respectively, 0.58, 0.50 and 0.70, confirming the limited discrimination ability of the model.

Predictions with the BN

In the task of discriminating between pregnancy and no-pregnancy, the BN model had accuracy equivalent to the decision tree, but a definitively higher AUC (0.72 versus 0.65) suggesting that the BN model can estimate more reliably the probability of pregnancy versus no pregnancy. As shown in Table II, the probabilities estimated by the model were well calibrated. For instance, the model correctly estimated a very low probability of pregnancy (<5%) in a set of 47 cycles in which only 3 clinical pregnancies resulted (6% observed pregnancy rate).

Table II Calibration of probabilities for discrimination of pregnancy and no pregnancy

Predicted probability of pregnancy (%)	0–5	5–10	10–20	20–30	30–40	40–60
No. of cycles	47	60	105	78	68	30
Observed pregnancy rate (%)	6	9	10	25	33	49
95% CI of the pregnancy rate	0–13	2–16	4–16	15–35	22–44	31–67

The conditions under which the model estimated pregnancy to be more probable than no pregnancy were (i) for IVF, the presence of two or more top-history embryos and age <34 years and (ii) for ICSI, the presence of 3 top-history embryos and age <34 years. Such conditions were quite coherent with those identified by the decision tree, and further confirmed the importance of top-history embryos and the woman's age for the establishment of a pregnancy.

However, the BN model provided an important improvement over the decision tree in discriminating between no pregnancy, single pregnancy and twin pregnancy, as indicated by the AUC measured in cross-validation, and corresponding to 0.72, 0.64 and 0.83, respectively. A crosscheck of the observed and predicted rates of clinical pregnancy for the cohort is given in Table III.

Additional improvement of the model performance was attempted by introducing further explaining variables. More specifically, maternal receptivity was also made to depend on the occurrence of previous pregnancies, and embryo viability was also made to depend on the use of sperm of proven fertility. Such attempts, despite potentially yielding a more detailed representation of the IVF-ICSI cycles, failed to improve the model's predictivity suggesting that a more complex model is not necessarily a better predictor, especially if the data set contains partially observed variables.

The estimates of maternal receptivity and embryo viability are shown in Tables IV and V, respectively. These probabilities cannot be trivially estimated from data, since for cycles resulting in no pregnancy it is unknown whether the mother was not receptive or the embryos were not viable. Therefore, it was necessary to adopt the Expectation–Maximization algorithm to estimate such probabilities. Age appeared to have a bigger impact on maternal receptivity than the number of previous IVF cycles (Table IV). Regarding embryo viability (Table V), there was an estimated difference of ~15 points between the viability of top-history and top-quality embryos, and between the viability of top and non-top embryos. Moreover, the estimated viability decreased ~7 points after ICSI, compared with IVF. The estimates showed that embryo score had a bigger impact on viability than the insemination technique.

Using expected utility

The elaboration of the described estimates (age, previous cycles, embryo quality and insemination technique) permitted the set-up of a prognostic model to be used prospectively when deciding the number and type of embryos to be transferred.

Table III Observed and predicted rates of clinical pregnancy for cycles transferred with different number of embryos, ranging from 1 to 3

Transferred embryos	Total cycles	Implanted embryos					
		0		1		2	
		Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)
1	108	93	90	7	10	—	—
2	179	76	80	18	18	6	2
3	101	73	73	21	20	6	5

Table IV Probability of the mother being receptive, as a function of the maternal age and the number of previous cycles (95% confidence intervals in parentheses)

Age (years)	Previous cycles	
	0–2	>2
<34	0.73 (0.63–0.83)	0.69 (0.59–0.79)
34–40	0.60 (0.54–0.66)	0.55 (0.49–0.61)
>40	0.41 (0.31–0.51)	0.19 (0.09–0.29)

Table V Probability of embryo implantation (assuming the mother to be receptive), as a function of the score and of the technique (95% confidence interval in parenthesis)

Embryo score	Technique	
	IVF	ICSI
Non-top	0.11 (0.08–0.14)	0.07 (0.04–0.10)
Top quality	0.25 (0.18–0.32)	0.22 (0.15–0.29)
Top history	0.48 (0.41–0.55)	0.34 (0.27–0.41)

Figure 3 shows, as an example, the expected utilities for different transfer possibilities, based on the probabilities computed by the BN model for a woman with age <34 years. Each row corresponds to a specific combination of her body build, insemination technique and number of previous cycles; once the appropriate row has been selected, the expected utility of transferring a different set of embryos can be read. A white background highlights the set of embryos with the highest expected utility. For example, if three top-history embryos (H) are available, the highest expected utility is attained in most situations by transferring only two of them because of the high risk of multiples. However in the case of overweight and ICSI, transferring all three embryos attains the maximum utility. Conversely, in the case of underweight women, the maximum utility is derived from the transfer of only one embryo (compare the columns 'HHH', 'HH' and 'H'). If three non-top embryos (*n*) are available, the highest utility is attained by transferring three embryos in all situations (compare the columns 'nnn', 'nn' and 'n'). Finally, if two top-quality (*t*) and one non-top embryos are available, the highest utility is obtained by transferring the three of them, apart from the case of underweight women, where the maximum utility is associated with the transfer of two top-quality embryos (compare columns 'ttn' and 'tt').

Probabilities of pregnancy calculated before and after embryo culture

The BN model can be used to provide information on the chances of pregnancy for couples before they start the treatment. In this situation, the quality of the resulting embryos is obviously unknown.

For example, in the case of a couple where the woman is <34 years old at her first stimulation and the male partner has normal semen indices, the BN model would estimate the probability of having 0, 1, 2 or 3 gestational sacs as, respectively, 69, 26, 5 and ~0%, with a probability of pregnancy corresponding to 31% (26 + 5), compared with the 20% which is the average in the studied sample.

If the information about embryo quality would be added, it becomes evident how the prediction of the probability of pregnancy varies as additional information is collected, and the estimates would be different. Actually, if three top-history embryos would be transferred to the woman described above, the probability of having 0, 1, 2 or 3 gestational sacs would change to 37, 28, 27 and 8%; in the case of two top-history embryos being transferred the probabilities would be 46, 36, 17 and 0%.

Discussion

Since the hallmark of excellence in reproductive medicine is a consistently high pregnancy rate associated with a single implantation, the shift towards single-embryo transfer should be adopted at least in couples at risk of multiple gestations. This policy implies having a reliable method of embryo selection that should be accurate enough as to identify those cases which, based on both embryo and patient characteristics, would possibly be penalized by single-embryo transfer. In this respect, innovative approaches have been proposed to improve the embryo selection process, some of which are invasive (Geraedts *et al.*, 2011; Schoolcraft *et al.*, 2011), others rely on extended culture systems (which are now questioned for the neonatal outcome of the babies born; Källén *et al.*, 2010), and others are dependent on sophisticated devices (Swain and Smith, 2011; Meseguer *et al.*, 2012). The common denominator for all these innovations is the need for validation, as their real efficacy is still largely unproven.

With these concepts in mind, the goal of this study was to design a system that, by combining embryo morphology and the clinical parameters that were found to be relevant in determining implantation, could predict the occurrence of a pregnancy and, more specifically, the chances of single, twin and multiple implantations. The proposed model can easily assist professionals in taking decisions at the time of embryo transfer by means of an objective statistical evaluation that only requires the input of specific data in a pre-set scheme. It is undisputable that, if the correct decision were taken, the chances of pregnancy would be maximized with a concomitant reduction in the incidence of multiple gestations, which is a priority in the field of reproductive medicine.

In consideration of the multifactorial typology of implantation where several variables play a role, designing a prediction model is a complex task. The approach that we adopted was based on the elaboration of previously generated data with known clinical outcomes in order to define the features relating patients' characteristics and embryo morphology, which were determinant in establishing a pregnancy. In this way a model was elaborated, based on cross-validation of data that could be operationally used in every setting to predict not only the event of pregnancy versus no pregnancy, but also the occurrence of single implantation versus twin or multiple gestational sacs.

As a first approach for the prediction of pregnancy, we used a decision tree by which, in agreement with other studies (Leushuis *et al.*, 2009), the accuracy was significantly higher than the trivial classifier

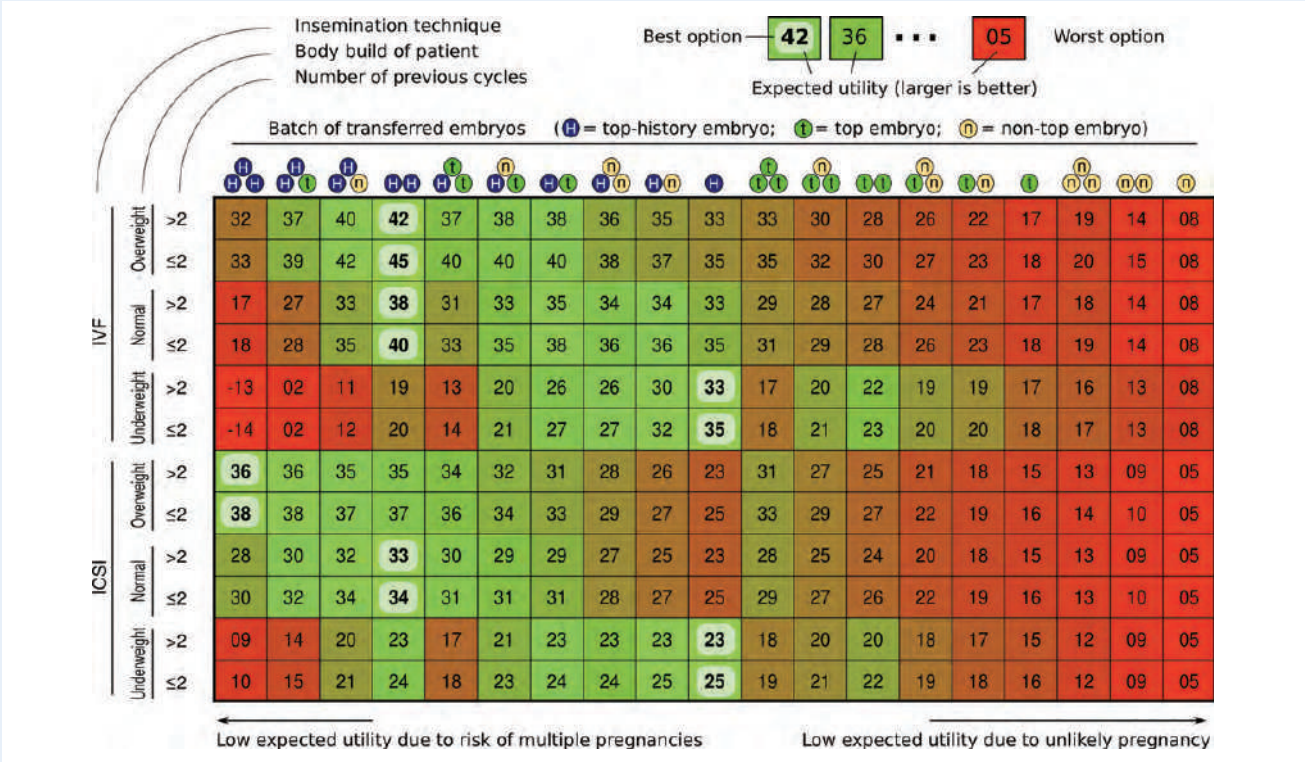

Figure 3 Expected utilities of different transfer possibility. The table is constructed for a patient <34 years old, considering all possible embryo type combinations (H = top history, t = top-quality, n = non top). Each row represents a specific combination of body build (underweight, normal and overweight) of the woman, insemination technique (IVF or ICSI) and number of previous cycles. The lighter the green colour, the higher the expected utility; the stronger the red colour, the lower the expected utility. For each row, a white background highlights the configuration with the highest expected utility.

(81.5 versus 80.2%, $P < 0.05$), but at such a narrow level that it was actually useless in practical terms. In addition, the model had poor performance in the more complicated task of discriminating between no pregnancy, pregnancy and multiple pregnancy. To formulate a model that could provide a solid support to IVF professionals, a BN model was proposed that made pregnancy to depend on the features selected as relevant by the training data, namely the age of the woman, the number of her previous IVF-ICSI cycles, the insemination technique and the score of the embryos. This model was also designed to assess how the probability of no pregnancy, single pregnancy and multiple pregnancy varies depending on the number and the type of the transferred embryos.

The relevance of embryo quality in establishing a pregnancy was immediately recognized by both statistical models, the decision tree and the BN. More specifically, it was clear that the probability of pregnancy strongly depended on the number of top-history embryos. In agreement with previous studies (Lan *et al.*, 2003), it was recognized that embryos that were top quality throughout all observations were those with the highest chances of implantation, confirming that synchrony of development plays a crucial role in determining viability (Lundin *et al.*, 2001; Lawler *et al.*, 2007; Magli *et al.*, 2007; Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). We set an exception to this rule for day 4 and day 5 embryos, for which the grade of top history was given even if they had been graded as top quality at all stages but one, provided that the single non-top assessment did not occur on the day of transfer. These observations suggest that later stages of embryo growth had higher sensitivity and specificity in the prediction of implantation (Rehman *et al.*, 2007), giving to the assessment of blastocyst morphology an extra value at the time of embryo selection. Interestingly enough, the same characteristics seem to be presented by day 4 embryos (Feil *et al.*, 2008) confirming that overcoming the trauma related to the activation of the embryonic genome is a sign of very good prognosis.

As expected, the concomitant analysis of the other variables indicated the negative association of implantation with age and the number of previous cycles, while a normal sperm sample was positively associated with a favourable outcome (Table V), suggesting that the capacity of activating the oocyte and guiding the first cleavage divisions are crucial events in determining the embryo fate (Heytens *et al.*, 2009; Speyer *et al.*, 2010). For instance, in the last couple of years, many contributions have demonstrated the relevance of sperm indices in affecting implantation due to several defects which are more common in pathological samples (Magli *et al.*, 2009; Speyer *et al.*, 2010; Ajduk *et al.*, 2011; Hammoud *et al.*, 2011).

Altogether these observations lead to the conclusion that, of the variables examined, age, number of previous cycles, insemination technique and embryo quality are the crucial factors affecting maternal

receptivity and embryo viability. The derived estimates calculated by the BN model were reasonably coherent with those made by Roberts *et al*., who reported an average of 0.36 for receptivity and of 0.32 for viability (Roberts *et al*., 2009).

As an additional advantage, the BN model proposed here was able to discriminate with reasonable reliability among no pregnancy, single pregnancy and twin pregnancy, although the number of twin pregnancies included in the data set was only 17. The corresponding estimates depended on the concept of utility that describes the most desirable clinical outcome for each couple. The weight of utility can be easily modified by IVF professionals at each calculation by keeping into consideration the negative effects related to multiple pregnancies especially in consideration of the woman's physical condition (Fig. 3). We made it depend on her body build, but additional factors could be also included, as the model is extremely flexible.

The benefit of a model that can support the decision at the time of transfer would be especially evident in case of having several embryos in culture. The scheme proposed here was based on three embryos that by law could be left in culture, and this notably decreased the number of observations and the possibility of selection. However, the BN model can be straightforwardly expanded to manage a larger number of embryos; in this case, the calculation of the expected utilities could be of even greater advantage in supporting decisions by offering wide ranges of possible combinations per each cycle. This would be especially important in those situations where the maternal receptivity and the morphological quality of the embryos suggest the transfer of more than one embryo. It is actually clear that the policy of single-embryo transfer is not the best option for all cases, and having an objective support at this stage would be of clear benefit (McLernon *et al*., 2010; Jonsdottir *et al*., 2011).

Should specific morphological and developmental details coming from morphokinetic studies be confirmed to be crucial for implantation (Wong *et al*., 2010; Meseguer *et al*., 2011), the corresponding features could be included to make the predictor even stronger.

As a further result, the BN model was able to identify cycles having an extremely low chance of implantation (Table II). This opens up to a series of considerations regarding the rationale of cancelling transfers with a very negative prognosis, avoiding, for the patient, the stress of a hopeless transfer with a minimal loss of pregnancies. Although this may be an acceptable approach from a statistical point of view that actually permits verification of the fitness of the utility model, from an ethical perspective it poses great difficulties especially related to the fact that embryos have already been generated.

In conclusion, the present study showed that all the prerequisites are given to construct a model aimed at supporting decisions and making predictions that are specific for every single case. The derived figures are clinically reasonable and have the advantage of being reproducible and objective, as confirmed by several measurements like AUC and the calibration probabilities. The preliminary data derived from its prospective application at the time of transfer (39 cycles so far) confirm its clinical validity resulting in 12 clinical pregnancies (31% clinical pregnancy rate): all singletons (in two cases three embryos were transferred) and one twin (two embryos were transferred). Moreover, the model is flexible enough to make predictions on the chances of pregnancy even before the couple starts a treatment cycle resulting in great acceptance from couples, although the prediction is more robust when the information on embryo quality is added (Smeenk *et al*., 2000). It is clear that some utility profiles can vary depending on each center's policy and experience, and they can be quickly changed by the IVF professionals to personalize the model represented in Fig. 3. In this way, it is possible to calculate for each couple the expected utility regarding which embryo and how many embryos to transfer for the most desired clinical outcome.

Authors' roles

L.G. formulated the study design, participated in manuscript drafting and critical discussion and approved the final version; M.C.M. participated in the study design, execution, analysis, manuscript drafting and critical discussion and approved the final version; A.G. participated in the study execution and critical discussion and approved the final version; C.G. participated in the study execution and critical discussion and approved the final version; L.M.G. participated in the study design, analysis and critical discussion and approved the final version; G.C. designed and validated the predictive models, participated in the study design, execution, analysis, manuscript drafting and critical discussion and approved the final version.

Funding

This study was supported by a grant, CTI Medtech Project Number: 9707.1 PFLS-L, Swiss Confederation.

Conflict of interest

None declared.

References

Ajduk A, Ilozue T, Windsor S, Yu Y, Seres KB, Bomphrey RJ, Tom BD, Swann K, Thomas A, Graham C *et al*. Rhythmic actomyosin-driven contractions induced by sperm entry predict mammalian embryo viability. *Nat Commun* 2011;**2**:417.

Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod* 2011; **26**:1270–1283.

Beuchat A, Thevenaz P, Unser M, Ebner T, Senn A, Urner F, Germond M, Sorzano CO. Quantitative morphometrical characterization of human pronuclear zygotes. *Hum Reprod* 2008;**23**:1983–1992.

Brezinova J, Oborna I, Svobodova M, Fingerova H. Evaluation of day one embryo quality and IVF outcome—a comparison of two scoring systems. *Reprod Biol Endocrinol* 2009;**7**:9.

Brison DR, Hollywood K, Arnesen R, Goodacre R. Predicting human embryo viability: the road to non-invasive analysis of the secretome using metabolic footprinting. *Reprod Biomed Online* 2007;**15**:296–302.

De Placido G, Wilding M, Strina I, Alviggi E, Alviggi C, Mollo A, Varicchio M, Tolino A, Schiattarella C, Dale B. High outcome predictability after IVF using a combined score for zygote and embryo morphology and growth rate. *Hum Reprod* 2002;**17**:2402–2409.

Dumoulin JC, Land JA, Van Montfoort AP, Nelissen EC, Coonen E, Derhaag JG, Schreurs IL, Dunselman GA, Kester AD. Effect of in vitro culture of human embryos on birthweight of newborns. *Hum Reprod* 2010;**25**:605–612.

Feil D, Henshaw RC, Lane M. Day 4 embryo selection is equal to Day 5 using a new embryo scoring system validated in single embryo transfers. *Hum Reprod* 2008;**7**:1505–1510.

Ferraretti AP, Gianaroli L, Magli MC, D'Angelo A, Farfalli V, Montanaro N. Exogenous LH in COH for ART: when and which? *Fertil Steril* 2004; **82**:1521–1526.

Gelbaya TA, Tsoumpou I, Nardo LG. The likelihood of live birth and multiple birth after single versus double embryo transfer at the cleavage stage: a systematic review and meta-analysis. *Fertil Steril* 2010;**94**:936–945.

Geraedts J, Montag M, Magli MC, Repping S, Handyside A, Staessen C, Harper J, Schmutzler A, Collins J, Goossens V *et al.* Polar body array CGH for prediction of the status of the corresponding oocyte. Part I: clinical results. *Hum Reprod* 2011;**26**:3173–3180.

Gianaroli L, Magli MC, Ferraretti AP, Fortini D, Grieco N. Pronuclear morphology and chromosomal abnormalities as scoring criteria for embryo selection. *Fertil Steril* 2003;**80**:341–349.

Gianaroli L, Magli MC, Ferraretti AP. Sperm and blastomere aneuploidy detection in reproductive genetics and medicine. *J Histochem Cytochem* 2005a;**53**:261–268.

Gianaroli L, Gordts S, D'Angelo A, Magli MC, Brosens I, Cetera C, Campo R, Ferraretti AP. Effect of the inner myometrium fibroid on the reproductive outcome after IVF. *Reprod Biomed Online* 2005b; **10**:473–477.

Gianaroli L, Magli MC, Ferraretti AP, Lappi M, Borghi E, Erimini B. Oocyte euploidy, pronuclear zygote morphology and embryo chromosomal complement. *Hum Reprod* 2007;**22**:241–249.

Gianaroli L, Magli MC, Cavallini G, Crippa A, Capoti A, Resta S, Robles F, Ferraretti AP. Predicting aneuploidy in human oocytes: key factors which affect the meiotic process. *Hum Reprod* 2010;**25**:2374–2386.

Guzeloglu-Kayisli O, Kayisli UA, Taylor HS. The role of growth factors and cytokines during implantation: endocrine and paracrine interactions. *Semin Reprod Med* 2009;**27**:62–79.

Hammoud SS, Nix DA, Hammoud AO, Gibson M, Cairns BR, Carrell DT. Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Hum Reprod* 2011;**26**:2558–2569.

Heytens E, Parrington J, Coward K, Young C, Lambrecht S, Yoon S-Y, Fissore RA, Hamer R, Deane CM, Ruas M *et al.* Reduced amounts and abnormal forms of phospholipase C zeta (PLCζ) in spermatozoa from infertile men. *Hum Reprod* 2009;**24**:2417–2428.

Horsthemke B, Ludwig M. Assisted reproduction: the epigenetic perspective. *Hum Reprod Update* 2005;**11**:473–482.

Jones GM, Trounson AO, Vella PJ, Thouas GA, Lolatgis N, Wood C. Glucose metabolism of human morula and blastocyst-stage embryos and its relationship to viability after transfer. *Reprod Biomed Online* 2001;**3**:124–132.

Jones GM, Cram DS, Song B, Magli MC, Gianaroli L, Lacham-Kaplan O, Findlay JK, Jenkin G, Trounson AO. Gene expression profiling of human oocytes following in vivo or in vitro maturation. *Hum Reprod* 2008a;**23**:1138–1144.

Jones G, Cram DS, Song B, Kokkali G, Pantos K, Trounson AO. Novel strategy with potential to identify developmentally competent IVF blastocysts. *Hum Reprod* 2008b;**23**:1748–1759.

Jonsdottir I, Lundin K, Bergh C. Double embryo transfer gives good pregnancy and live birth rates in poor responders with a modest increase in multiple birth rates: results from an observational study. *Acta Obstet Gynecol Scand* 2011;**90**:761–766.

Källén B, Finnström O, Lindam A, Nilsson E, Nygren KG, Olausson PO. Blastocyst versus cleavage stage transfer in in vitro fertilization: differences in neonatal outcome? *Fertil Steril* 2010;**94**: 1680–1683.

Katz-Jaffe MG, Gardner DK, Schoolcraft WB. Proteomic analysis of individual human embryos to identify novel biomarkers of development and viability. *Fertil Steril* 2006;**85**:101–107.

Koller D, Friedman N. *Probabilistic Graphical Models*. MIT Press, Cambridge, Massachusetts London, England 2009.

Kovacic B, Vlaisavljevic V, Reljic M, Cizek-Sajko M. Developmental capacity of different morphological types of day 5 human morulae and blastocysts. *Reprod Biomed Online* 2004;**8**:687–694.

Kuliev A, Cieslak J, Verlinsky Y. Frequency and distribution of chromosome abnormalities in human oocytes. *Cytogenet Genome Res* 2005; **111**:193–198.

Lan K, Huang F, Lin Y, Kung F, Hsieh C, Huang H, Tan PH, Chang SY. The predictive value of using a combined Z-score and day 3 embryo morphology score in the assessment of embryo survival on day 5. *Hum Reprod* 2003;**18**:1299–1306.

Lawler C, Baker HWG, Edgar DH. Relationships between timing of syngamy, female age and implantation potential in human in vitro-fertilised oocytes. *Reprod Fertil Dev* 2007;**19**:482–487.

Leushuis E, Van Der Steeg J, Steures PB, Eijkemans M, Van Der Veen F, Mol BW, Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**15**:537–552.

Loi K, Prasath EB, Huang ZW, Loh SF, Loh SKE. A cumulative embryo scoring system for the prediction of pregnancy outcome following intracytoplasmic sperm injection. *Singapore Med J* 2008;**49**:221–227.

Lundin K, Bergh C, Hardarson T. Early embryo cleavage is a strong indicator of embryo quality in human IVF. *Hum Reprod* 2001;**16**:2652–2657.

Magli MC, Gianaroli L, Ferraretti AP, Lappi M, Ruberti A, Farfalli V. Embryo morphology and development is dependent on the chromosomal complement. *Fertil Steril* 2007;**87**:534–541.

Magli MC, Gianaroli L, Ferraretti AP, Gordts S, Fredericks V, Crippa A. Paternal contribution to aneuploidy in preimplantation embryos. *Reprod Biomed Online* 2009;**18**:536–542.

Margalioth EJ, Ben-Chetrit A, Gal M, Eldar-Geva T. Investigation and treatment of repeat implantation failure following IVF-ET. *Hum Reprod* 2006;**21**:3036–3043.

McLernon DJ, Harrild K, Bergh C, Davies MJ, de Neubourg D, Dumoulin JC, Gerris J, Kremer JA, Martikainen H, Mol BW *et al.* Clinical effectiveness of elective single versus double embryo transfer: meta-analysis of individual patient data from randomised trials. *BMJ* 2010;**341**:c6945.

Meseguer M, Herrero J, Tejera A, Hilligsøe KM, BirgerRamsing N, Remohı J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod* 2011;**26**:2658–2671.

Meseguer M, Kruhne U, Laursen S. Full in vitro fertilization laboratory mechanization: toward robotic assisted reproduction? *Fertil Steril* 2012; **97**:1277–1286.

Munné S, Chen S, Colls P, Garrisi J, Zheng X, Cekleniak N, Lenzi M, Hughes M, Fischer J, Garrisi M *et al.* Maternal age, morphology, development and chromosome abnormalities in over 6000 cleavage-stage embryos. *Reprod Biomed Online* 2007;**14**:628–634.

Norwitz ER, Schust DJ, Fisher SJ. Implantation and the survival of early pregnancy. *N Engl J Med* 2001;**345**:1400–1408.

Pandian Z, Bhattacharya S, Ozturk O, Serour G, Templeton A. Number of embryos for transfer following in-vitro fertilisation or intra-cytoplasmic sperm injection. *Cochrane Database Syst Rev* 2009;**2**:CD003416.

Racowsky C, Stern JE, Gibbons WE, Behr B, Pomeroy KO, Biggers JD. 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* 2011;**95**:1985–1989.

Rehman KS, Bukulmez O, Langley M, Carr BR, Nackley AC, Doody KM, Doody KJ. Late stages of embryo progression are a much better

predictor of clinical pregnancy than early cleavage in intracytoplasmic sperm injection and in vitro fertilization cycles with blastocyst-stage transfer. *Fertil Steril* 2007;**87**:1041–1052.

Roberts S. Models for assisted conception data with embryo-specific covariates. *Stat Med* 2007;**26**:156–170.

Roberts S, Fitzgerald C, Brison D. Modelling the impact of single embryo transfer in a national health service IVF programme. *Hum Reprod* 2009; **24**:122–131.

Roberts SA, Hirst WM, Brison DR, Vail A. Embryo and uterine influences on IVF outcomes: an analysis of a UK multi-centre cohort. *Hum Reprod* 2010;**25**:2792–2802.

Saith R, Srinivasan A, Michie D, Sargent IL. Relationships between the developmental potential of human in-vitro fertilization embryos and features describing the embryo, oocyte and follicle. *Hum Reprod Update* 1998;**4**:121–134.

Santos M, Kuijk EW, Macklon NS. The impact of ovarian stimulation for IVF on the developing embryo. *Reprod* 2010;**139**:23–34.

Schoolcraft WB, Treff NR, Stevens JM, Ferry K, Katz-Jaffe M, Scott RT Jr. Live birth outcome with trophectoderm biopsy, blastocyst vitrification, and single-nucleotide polymorphism microarray-based comprehensive chromosome screening in infertile patients. *Fertil Steril* 2011; **96**:638–640.

Simon C, Martin JC, Pellicer A. Paracrine regulators of implantation. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**14**:815–826.

Sjöblom P, Menezes J, Cummins L, Mathiyalagan B, Costello MF. Prediction of embryo developmental potential and pregnancy based on early stage morphological characteristics. *Fertil Steril* 2006;**86**:848–861.

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

Speyer BE, Pizzey AR, Ranieri M, Joshi R, Delhanty JD, Serhal P. Fall in implantation rates following ICSI with sperm with high DNA fragmentation. *Hum Reprod* 2010;**25**:1609–1618.

Swain JE, Smith GD. Advances in embryo culture platforms: novel approaches to improve preimplantation embryo development through modifications of the microenvironment. *Hum Reprod Update* 2011; **17**:541–557.

Volpes A, Sammartano F, Coffaro F, Mistretta V, Scaglione P, Allegra A. Number of good quality embryos on day 3 is predictive for both pregnancy and implantation rates in in vitro fertilization/ intracytoplasmic sperm injection cycles. *Fertil Steril* 2004;**82**:1330–1336.

Witten IH, Frank E, Hal M. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd edn. San Francisco, CA: Morgan Kaufmann Publishers, 2011.

Wong CC, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, Reijo Pera RA. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol* 2010;**28**:1115–1121.

World Health Organization. Obesity: preventing and managing the global epidemic. Report of a WHO Consultation on Obesity. WHO Technical Report Series 894. Geneva, Switzerland: World Health Organization, 2000.

Zhou H, Weinberg C. Evaluating effects of exposures on embryo viability and uterine receptivity in in vitro fertilization. *Stat Med* 1998; **17**:1601–1612.