Which set of embryo variables is most predictive for live birth? A prospective study in 6252 single embryo transfers to construct an embryo score for the ranking and selection of embryos

A. Rhenman1,*, L. Berglund2, T. Brodin3, M. Olovsson1, K. Milton3, N. Hadziosmanovic2, and J. Holte1,3,4

1Department of Women’s and Children’s Health, Uppsala University, S-751 85 Uppsala, Sweden 2Uppsala Clinical Research Center, Akademiska sjukhuset (UCR), S-751 85 Uppsala, Sweden 3Carl von Linné Clinic, Uppsala Science Park, S-751 83 Uppsala, Sweden 4Centre for Reproductive Biology in Uppsala (CRU), PO Box 7054, S-750 07 Uppsala, Sweden

*Correspondence address. E-mail: axelrhenman@yahoo.se

Submitted on February 5, 2014; resubmitted on October 9, 2014; accepted on October 15, 2014

STUDY QUESTION: Which embryo score variables are most powerful for predicting live birth after single embryo transfer (SET) at the early cleavage stage?

SUMMARY ANSWER: This large prospective study of visual embryo scoring variables shows that blastomere number (BL), the proportion of mononucleated blastomeres (NU) and the degree of fragmentation (FR) have independent prognostic power to predict live birth.

WHAT IS KNOWN ALREADY: Other studies suggest prognostic power, at least univariately and for implantation potential, for all five variables. A previous study from the same centre on double embryo transfers with implantation as the end-point resulted in the integrated morphology cleavage (IMC) score, which incorporates BL, NU and EQ.

STUDY DESIGN, SIZE AND DURATION: A prospective cohort study of IVF/ICSI SET on Day 2 (n = 6252) during a 6-year period (2006–2012). The five variables (BL NU, FR and symmetry of cleavage (SY)) were scored in 3- to 5-step scales and subsequently related to clinical pregnancy and LBR.

PARTICIPANTS/MATERIALS, SETTING, METHODS: A total of 4304 women undergoing IVF/ICSI in a university-affiliated private fertility clinic were included. Generalized estimating equation models evaluated live birth (yes/no) as primary outcome using the embryo variables as predictors. Odds ratios with 95% confidence intervals and P-values were presented for each predictor. The C statistic (i.e. area under receiver operating characteristic curve) was calculated for each model. Model calibration was assessed with the Hosmer–Lemeshow test. A shrinkage method was applied to remove bias in C statistics due to over-fitting.

MAIN RESULTS AND THE ROLE OF CHANCE: LBR was 27.1% (1693/6252). BL, NU, FR and EQ were univariately highly significantly associated with LBR. In a multivariate model, BL, NU and FR were independently significant, with C statistic 0.579 (age-adjusted C statistic 0.637). EQ did not retain significance in the multivariate model. Prediction model calibration was good for both pregnancy and live birth. We present a ranking tree with combinations of values of the BL, NU and FR embryo variables for optimal selection of the embryo/s to transfer, providing a revised IMC score. The five embryo variables had similar effects over all age groups.

LIMITATIONS, REASONS FOR CAUTION: Limitations of the present study are those inherent for real-time visual scoring, including risks of inter-observer variation and the hazards of fixed time-point scoring procedures in a dynamic process. The study is restricted to Day-2 transfers.

WIDER IMPLICATIONS OF THE FINDINGS: To our knowledge this is the largest prospective, SET study performed with the explicit aim of constructing an evidence-based embryo score for the ranking and selection of early cleavage stage embryos. In line with previous research, our data suggest that the symmetry of cleavage variable may be omitted when scoring embryos in the early cleavage stage. We suggest that, following validation in other populations, the revised IMC score may be used when international standards for embryo scoring are discussed.
Introduction

Embryo quality is widely regarded as the single most important factor for the chance of a successful IVF treatment (Harton et al., 2013). Scoring of embryo morphology and cleavage rate is a clinically useful method in embryo selection and may serve as a proxy for embryo quality (Gardner et al., 2000, 2004; Balaban et al., 2006; Holte et al., 2007).

There is still a lack of evidence-based standardized morphological and cleavage variables for ranking and selection of the embryo with the greatest chance of resulting in a live baby. The variation in embryo scoring models makes comparisons between different clinics and treatment strategies difficult, a fact which is highlighted when accumulating national data (Vernon et al., 2011). Solid data on how to rank and select individual embryos with different types and degrees of deviations from the ‘ideal’ morphologic and cleavage appearance are still scarce. Thus, in clinical practice, it is not clear if, for example, an embryo with ‘too fast’ or ‘too slow’ cleavage and no fragmentation should be selected for transfer before an embryo with normal cleavage rate and moderate fragmentation, or if a slight variation in blastomere size is a worse prognostic sign than the absence of a visible normal nucleus in a proportion of the blastomeres.

In a previous study it was investigated whether the five commonly used embryo variables, namely blastomere number (BL), the proportion of mononucleated blastomeres (or presence of multinuclearity; NU), the degree of fragmentation (FR), the size variation of the blastomeres (equality, EQ) and the symmetry of the embryo (SY), were independent predictors of implantation potential of the embryo (Holte et al., 2007). Of those five variables, only three were independent predictors in multivariate regression analysis: BL, EQ and NU. Although also significant in univariate testing, the variables FR and SY did not retain significance in the multivariate model. The Holte et al. (2007) study was conducted on double embryo transfers (DET), and thus the primary outcome was either the implantation of both or neither of the transferred embryos. The study resulted in a clinically useful algorithm, the integrated morphology cleavage score (IMC), which enables evidence-based scoring of embryos on Day 2 after oocyte retrieval in a 10-point scale. The IMC score forms an essential part of a validated prediction model as an aid in the decision between SET and DET (Holte et al., 2011a), a model which has been in clinical use in several clinics for a number of years.

The aim of the present study was to re-evaluate the same five embryo variables in a second prospective study, using single embryo transfer (SET) only and with live birth rate (LBR) as primary end-point.

Materials and Methods

All data were collected prospectively. Unselected treatments at the Carl von Linné Clinic during a period of 6 years (2006–2012) were included. Only treatments that resulted in a SET on Day 2 after oocyte retrieval were studied regardless of whether SET was elective or not. Patients (i.e. treatments) were enrolled regardless of cause and duration of infertility and regardless of infertility being primary or secondary.

Ovarian stimulations were conducted with individual doses of recombinant FSH in the form of follitropin alfa (Merck Serono, Darmstadt, Germany) or follitropin beta (MSD, Whitehouse Station, New Jersey) in 79% of the cycles and with human menopausal gonadotrophin (hMG) (Ferring, Lausanne, Switzerland) in 21% of the cycles. In 90% of the cycles, patients underwent long GnRH-agonist down-regulation with nafarelin (Pfizer, New York, New York) or buserelin (Sanofi, Paris, France). In 10% of the treatments, a GnRH antagonist (Cetrorelix, Merck Serono, Darmstadt, Germany; or Ganirelix, MSD, Whitehouse Station, New Jersey) was used, starting on the fifth day of stimulation. Final oocyte maturation was triggered with hCG using either 10 000 IU of urine derived hCG (MSD, Whitehouse Station, New Jersey) or 250 µg (6500 IU) of recombinant hCG (Merck Serono, Darmstadt, Germany), and oocytes were aspired after 35–37 h. Oocytes were insemminated (or injected with sperm after denudation for ICSI) after 2–6 h of incubation and then cultured in one of three culture mediums (MediCult, Denmark; Cook Medical, Australia or Vitrolife, Sweden) in a 5% CO2 incubator at 37°C. Fertilization was checked 16–20 h after the insemination. The embryos were evaluated 46–48 h after insemination by placing the Petri dish with the embryos on the heating plate under a stereomicroscope (Nikon SMZ 1500) at maximal magnification (×110). The embryos were gently rolled with a Pasteur pipette to assess the five morphological parameters. Light was directed obliquely through an angled mirror beneath the microscope to enable the visualization of the nuclei. The five embryo variables were scored as follows: BL was recorded as 2, 3, 4, 5 or ≥6, FR was scored as 0 (no fragments), 1 (≤10% fragmentation), 2 (>10 ≤25%), 3 (>25 ≤50%) and 4 (>50% fragmentation). EQ was scored according to a three-level system, where 0 denotes uniform size of the blastomeres, 1 denotes varying size but <50% variation and 2 denotes >50% variation in blastomere size. Similarly, SY was scored as 0, 1 and 2, describing full symmetry of the cleaved embryo, slightly asymmetric cleavage, and pronounced asymmetry, respectively. The difference between these two parameters is thus that EQ scores variation in blastomere size whereas SY scores the arrangement of the blastomeres three-dimensionally. The parameter NU was defined as the number of visible mononucleated blastomeres divided by the total number of blastomeres in the embryo (to correct for the cleavage rate); a ratio of 0–0.29 gave a NU score of 0, 0.30–0.69 gave a NU score of 1, 0.70–0.99 gave a score of 2, and a ratio of 1 equaled a NU score of 3. The NU score −1 denotes that the embryo contains at least one multinucleated blastomere. Supplementary Table SI gives all the possible NU scores for embryos with 2–8 blastomeres and varying proportions of visible nuclei. Supplementary Table SI is given to facilitate the NU scoring, which otherwise must be calculated manually for each combination. With similar definitions, these five variables were used in our previous study, which resulted in the IMC (Holte et al., 2007). The IMC included BL, NU and EQ, which remained significant after regression analysis, incorporated in an algorithm and presented as a 10-point scale. In that study, only DETs were included, and the end-point was clinical pregnancy. In the present study, the IMC was the basis for ranking and selection of embryos, but FR and SY were still scored for the purpose of comparing their potential to compete with the other three variables to be included in the new algorithm, which also predicts a slightly different end-point, LBR.

Embryos were transferred 51–55 h after oocyte retrieval. After embryo transfer, all patients were given luteal phase support vaginally for 2 weeks, either as 400 mg of micronized progesterone (vagitories, Apoteket AB, Stockholm, Sweden) or 100 mg vaginal tablets (Ferring, Lausanne,
Statistics

The sample size was based on the perceived length of the 95% confidence interval (CI) for the pregnancy odds ratio for the most important predictor; the absolute deviation of blastomeres from 4 (abs(BL-4)). This odds ratio was 0.39 in our previous study (Holte et al., 2007). We calculated that 5000 treatments would be sufficient to make the length of the 95% CI 0.2. Subsequently, as shown in the Results section, this power calculation proved valid, given the lengths of the calculated 95% CIs for odds ratios for abs(BL-4); the estimated odds ratio is 0.47 with 95% CI 0.39–0.58, which is highly informative.

The interrelations between the predictor variables were examined using Spearman’s rank correlation coefficients. Univariate relations between predictor variables and the outcomes pregnancy and live birth were described in figures and statistically analysed in generalized estimating equation (GEE) models (Zeger et al., 1988) with clinical pregnancy and live birth (yes/no) as outcome variables and embryo factors as predictors. GEE models take into account the dependence between repeated treatments for the same woman. Results were presented as odds ratios with 95% CIs and \( P \)-values for each predictor and c statistic, i.e. area under receiver operating characteristic curve, for each model.

A model’s predictive capacity, which is measured as the c statistic in this study, when estimated from the same data set as was used for model development, is commonly too optimistic compared with the accuracy found in new patients. This over-fitting is more likely to occur when the derivation data set is small, and/or there is a large number of candidate predictors compared with the number of outcomes (in our case the number of delivered babies) (Moons et al., 2004). Although none of these conditions characterized the present study, we still used a shrinkage method to adjust for possible over-fitting and remove bias in c statistics (Harrell et al., 1996). The method was applied to a logistic regression model with data from the first treatment for each woman.

Model calibration was assessed with the Hosmer–Lemeshow test (Hosmer and Lemeshow, 2000). The test assesses whether or not the observed event rates match expected event rates in subgroups based on predicted probabilities from the model. The Hosmer–Lemeshow test results are non-significant if the calibration is acceptable.

Each predictor’s (i.e. embryo variable’s) functional relation to the outcome was examined and, if necessary, the variable was transformed. Thereafter, a GEE multivariate regression model with forward selection of variables was applied to find the set of predictor variables with independent impact on the outcome. \( P \)-values from two-tailed tests \(< 5\%\) were considered statistically significant. From this model, estimated probabilities were calculated for each treatment using a weighed sum of mean embryo variable values where the estimated regression coefficients were used as weights. This calculated value is the estimated probability of pregnancy or a live birth. The relative merits of the predictor variables were illustrated by the calculation of the estimated probabilities (transferred to a scale 0–10, where 10 is the highest probability) for a number of combinations of the predictors’ values.

Ethical approval

The study was approved by the Regional Ethics Review Board, Uppsala, Sweden.

Written informed consent was obtained from all participants.

Results

The study cohort consisted of 6252 treatment cycles in 4304 couples. Table I shows the characteristics of the study cohort. Mean female age (SD) was 34 (4.2) years. The overall LBR was 27.1%.

Effects of embryo variables in univariate models

Figure 1 shows the relationships between each of the five embryo variables and predicted and observed LBRs. All variables except for SY were significantly associated with LBR. All variables except BL could best be described by linear functions; for BL, LBRs increased from 2- to 4-cell embryos but then declined again with higher numbers of blastomeres.

As BL was not linearly related to the outcome, it had to be de-composed into two variables. The linear BL (linBL) term is the number of blastomeres and the term abs(BL-4) is the deviation from the ideal number of blastomeres. Thus, the relation was best described by the combination of a linear term and the absolute deviation from 4 cells (e.g. 3 cells and 5 cells both have an absolute deviation of 1).

Table I Characteristics of the cohort in a prospective study of embryo variables and prediction of live birth.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, n</td>
<td>4304</td>
</tr>
<tr>
<td>Embryo transfers, n</td>
<td>6252</td>
</tr>
<tr>
<td>ET/woman, mean (min-max)</td>
<td>1.45 (1–7)</td>
</tr>
<tr>
<td>Couples with 1 treatment, n (%)</td>
<td>2884 (67.0)</td>
</tr>
<tr>
<td>Couples with 2 treatments, n (%)</td>
<td>1006 (23.4)</td>
</tr>
<tr>
<td>Couples with 3 treatments, n (%)</td>
<td>328 (7.6)</td>
</tr>
<tr>
<td>Couples with 4 or more treatments, n (%)</td>
<td>86 (2.0)</td>
</tr>
<tr>
<td>Female age in years, mean (SD)</td>
<td>34 (4.2)</td>
</tr>
<tr>
<td>Eggs per oocyte retrieval, mean (range)</td>
<td>9.75 (1–39)</td>
</tr>
<tr>
<td>Tubal factor, n (%)</td>
<td>692 (11.1)</td>
</tr>
<tr>
<td>Anovulation, n (%)</td>
<td>525 (8.4)</td>
</tr>
<tr>
<td>Unexplained infertility, n (%)</td>
<td>2512 (40.1)</td>
</tr>
<tr>
<td>Male factor, n (%)</td>
<td>1665 (26.6)</td>
</tr>
<tr>
<td>Other cause of infertility, n (%)</td>
<td>858 (13.8)</td>
</tr>
<tr>
<td>ICSI, n (%)</td>
<td>2164 (34.6)</td>
</tr>
<tr>
<td>IVF, n (%)</td>
<td>3445 (55.1)</td>
</tr>
<tr>
<td>Combined IVF/ICSI, n (%)</td>
<td>643 (10.3)</td>
</tr>
<tr>
<td>Pregnancy rate, %</td>
<td>34.3</td>
</tr>
<tr>
<td>Live birth rate, %</td>
<td>27.1</td>
</tr>
</tbody>
</table>

For category variables, n (%) is presented and for continuous variables, mean \pm SD or range (minimum; maximum) is presented. ET, embryo transfer.
The term $\text{abs(BL-4)}$ implies a symmetric relation but addition of the linear term introduces a slight asymmetric feature such that 5 cells gives a better outcome than 3 cells. Thus, the combination of these terms captures the asymmetric relation between BL and LBR evident from Fig. 1. In the following, BL refers to the combination of $\text{linBL}$ term and $\text{abs(BL-4)}$.

### Table II  Effects of single embryo variables on live birth rates: univariate analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OR (CI)</th>
<th>P-value</th>
<th>c-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{linBL}$</td>
<td>1.38 (1.15–1.65)</td>
<td>0.001</td>
<td>0.548</td>
</tr>
<tr>
<td>$\text{abs(BL-4)}$</td>
<td>0.39 (0.32–0.47)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>$\text{NU}$</td>
<td>1.68 (1.51–1.87)</td>
<td>&lt;0.0001</td>
<td>0.545</td>
</tr>
<tr>
<td>$\text{FR}$</td>
<td>0.77 (0.71–0.84)</td>
<td>&lt;0.0001</td>
<td>0.535</td>
</tr>
<tr>
<td>$\text{EQ}$</td>
<td>0.69 (0.61–0.78)</td>
<td>&lt;0.0001</td>
<td>0.537</td>
</tr>
<tr>
<td>$\text{SY}$</td>
<td>0.88 (0.75–1.03)</td>
<td>0.103</td>
<td>0.513</td>
</tr>
</tbody>
</table>

Odds ratios (OR) with 95% confidence intervals (CI) and P-values and c-statistics for the end-point live birth for each of the variables; Blastomere (BL), Nucleus (NU), Fragmentation (FR), Equality (EQ) and Symmetry (SY). Number of blastomeres is represented both by the linear number of blastomeres, $\text{linBL}$, and the absolute number of blastomeres – 4, $\text{abs(BL-4)}$, given the non-linear association between the blastomere variable and the end-point.

The term $\text{abs(BL-4)}$ implies a symmetric relation but addition of the linear term introduces a slight asymmetric feature such that 5 cells gives a better outcome than 3 cells. Thus, the combination of these terms captures the asymmetric relation between BL and LBR evident from Fig. 1. In the following, BL refers to the combination of $\text{linBL}$ term and $\text{abs(BL-4)}$.

Table II shows univariate odds ratios and the discriminative abilities for a live birth for the five embryo variables. BL had the highest discriminative capacity (c statistic 0.548), closely followed by NU and FR and EQ. Of the five embryo variables, only SY was univariately insignificant.

### Effects of embryo variables and their relative importance in multivariate models

In the GEE multivariate regression model, BL, NU and FR remained independent significant predictors for a live birth, but not EQ, which is why this variable was subsequently excluded. Table III shows the discriminative capacity for different combinations of the independently significant variables. The model including BL, NU and FR (Table III, model 1) had the highest discriminative capacity for prediction of live birth (c statistic 0.579). Removal of FR or NU from the multivariate model (model 2 and 3, respectively) resulted in a c statistic of 0.568 and 0.569, respectively. The combination NU and FR gave the lowest c statistic of 0.559 (model 4).

To construct the revised IMC score (rIMC) we transferred the resulting figure of predicted probabilities of live birth from the algorithm of model 1 (Table III) to a scale of 0–10. Supplementary Figure S1 displays all combinations of BL, NU and FR with corresponding rIMC scores, the value 10 (BL 4, NU 3, i.e. all four blastomeres with a visible nucleus, and FR 0) thus implying the highest probability of a live birth. Supplementary Table SI shows the calculation of all possible NU scores up to eight blastomeres and each possible rIMC score.

Examples of the effect of different rIMC scores: Transfer of an embryo with a score of 10 points resulted in a mean LBR of 33.1% (predicted: 33.2%; $N = 1126$), whereas the corresponding figure for a rIMC score of 9.1 was 29.3% (predicted: 30.3%; $N = 3469$). Table IV summarizes the observed and predicted LBRs for different levels of rIMC scores.

Age did not interact with any of the variables, i.e. the embryo variables had similar effects over all age groups. The age-adjusted c statistic for
were 0.593 and 0.593, respectively. and the c statistics were compared. The c statistics for these models and thereafter a shrinkage model using the same data was estimated model with data from the first treatment for each woman (Holte et al., 2007). We focused on evaluating five morphological variables at the early cleavage stage, determining each variable’s individual relationship with treatment outcome, and further aimed at determining the relative weight of these parameters to predict live birth. The study is unique in the sense that all studied embryos were scored prospectively and SET was used. We focused on evaluating five morphological variables at the early cleavage stage, determining each variable’s individual relationship with treatment outcome, and further aimed at determining the relative weight of these parameters to predict live birth. The study is a follow-up of our previous study, which investigated the same variables in DET and using implantation as the end-point (Holte et al., 2007). We found that the present embryo model, derived from only SET in a much larger cohort and with live birth as main end-point, is very similar to the one presented previously. In univariate analysis, the results were almost identical, confirming that treatment outcome is associated with cleavage rate in a non-linear way, and in a linear way with nuclear status, degree of fragmentation and variation in blastomere size. After regression modelling, the degree of fragmentation qualified in the final new model instead of the variation in blastomere size, which was incorporated in the previous model (Holte et al., 2007). Thus the new algorithm, the rIMC, incorporates cleavage rate, the percentage of blastomeres showing a normal nucleus, and the degree of fragmentation, whereas the variation in rIMC (0–10 point score) 0–4 4–5 5–8 8–9.9 10

<table>
<thead>
<tr>
<th>rIMC (0–10 point score)</th>
<th>0–4</th>
<th>4–5</th>
<th>5–8</th>
<th>8–9.9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations (n)</td>
<td>543</td>
<td>144</td>
<td>473</td>
<td>3966</td>
<td>1126</td>
</tr>
<tr>
<td>Observed live birth, n (%)</td>
<td>29</td>
<td>24</td>
<td>100</td>
<td>1167</td>
<td>373</td>
</tr>
<tr>
<td>Predicted live birth %</td>
<td>6.4</td>
<td>14.9</td>
<td>21.8</td>
<td>29.9</td>
<td>33.2</td>
</tr>
</tbody>
</table>

The cohort is divided into five revised Integrated Morphology Cleavage score (rIMC) levels: 0–4, 4–5, 5–8, 8–9.9 and 10. For each rIMC group the total number of observations and the observed and predicted live birth rates are displayed.

model 1 was 0.637. LBRs for different embryo scores were calculated in seven age groups and the resulting graphs are depicted in Fig. 2.

Adjustment for over-fitting

To adjust for over-fitting, we first estimated a logistic regression model with data from the first treatment for each woman (n = 4452) and thereafter a shrinkage model using the same data was estimated and the c statistics were compared. The c statistics for these models were 0.593 and 0.593, respectively.

Correlation of the variables

Overall, the embryo variables displayed modest intercorrelations (Table V). The strongest correlation was between abs(BL-4) and NU (r = −0.40), followed by NU and EQ (r = −0.35), abs(BL-4) and EQ (r = 0.25) and FR and NU (r = −0.23).

Calibration of the multivariate model

Figure 3 shows the association between the mean predicted probability of a pregnancy and the observed pregnancy rates. The corresponding results for LBRs are given in Fig. 4. The goodness of fit was good for both outcomes, assessed with the Hosmer–Lemeshow test, which was insignificant for both outcomes (P = 0.60 for the prediction of pregnancy and P = 0.39 for a live birth).

Discussion

This is by far the largest study on classic embryo morphological markers and their relation to chance of a successful IVF/ICSI treatment. The study is unique in the sense that all studied embryos were scored prospectively and SET was used. We focused on evaluating five morphological variables at the early cleavage stage, determining each variable’s individual relationship with treatment outcome, and further aimed at determining the relative weight of these parameters to predict live birth. The study is a follow-up of our previous study, which investigated the same variables in DET and using implantation as the end-point (Holte et al., 2007). We found that the present embryo model, derived from only SET in a much larger cohort and with live birth as main end-point, is very similar to the one presented previously. In univariate analysis, the results were almost identical, confirming that treatment outcome is associated with cleavage rate in a non-linear way, and in a linear way with nuclear status, degree of fragmentation and variation in blastomere size. After regression modelling, the degree of fragmentation qualified in the final new model instead of the variation in blastomere size, which was incorporated in the previous model (Holte et al., 2007). Thus the new algorithm, the rIMC, incorporates cleavage rate, the percentage of blastomeres showing a normal nucleus, and the degree of fragmentation, whereas the variation in

Figure 2 Mean live birth rates at revised integrated morphology cleavage (rIMC) scores from 1 to 10 points in different age groups. The impact of embryo score on implantation is proportionally affected by age over the entire range of embryo scores.

Figure 3 shows the association between the mean predicted probability of a pregnancy and the observed pregnancy rates. The corresponding results for LBRs are given in Fig. 4. The goodness of fit was good for both outcomes, assessed with the Hosmer–Lemeshow test, which was insignificant for both outcomes (P = 0.60 for the prediction of pregnancy and P = 0.39 for a live birth).

Discussion

This is by far the largest study on classic embryo morphological markers and their relation to chance of a successful IVF/ICSI treatment. The study is unique in the sense that all studied embryos were scored prospectively and SET was used. We focused on evaluating five morphological variables at the early cleavage stage, determining each variable’s individual relationship with treatment outcome, and further aimed at determining the relative weight of these parameters to predict live birth. The study is a follow-up of our previous study, which investigated the same variables in DET and using implantation as the end-point (Holte et al., 2007). We found that the present embryo model, derived from only SET in a much larger cohort and with live birth as main end-point, is very similar to the one presented previously. In univariate analysis, the results were almost identical, confirming that treatment outcome is associated with cleavage rate in a non-linear way, and in a linear way with nuclear status, degree of fragmentation and variation in blastomere size. After regression modelling, the degree of fragmentation qualified in the final new model instead of the variation in blastomere size, which was incorporated in the previous model (Holte et al., 2007). Thus the new algorithm, the rIMC, incorporates cleavage rate, the percentage of blastomeres showing a normal nucleus, and the degree of fragmentation, whereas the variation in
blastomere size and the symmetry of the cleavage did not qualify for the final model.

The use of poor-quality prediction models could have a negative effect on clinical strategies by introducing the illusion of objective improvement over clinical judgement (Leushuis et al., 2009). Many previous studies on embryo morphology have focused on a single factor, such as cleavage rate, degree of fragmentation, uneven cleavage and multinuclearity, all factors which are associated with implantation and birth rates in univariate analyses (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). There are, however, only a limited number of studies that take a multi-dimensional approach, which better reflects the common clinical situation, when ranking embryos with various degrees and types of morphological deviation is needed (Scott, 2003; Sjoblom et al., 2006; Guerif et al., 2007; Holte et al., 2007; Rehman et al., 2007; Racowsky et al., 2009; Ahlstrom et al., 2011; Fauque et al., 2013). A multivariate model is necessary to find out which variables have independent, and thus true predictive, impact on the outcome. Some variables may lack biological significance, but show an association with other variables of real importance and thus falsely give the impression of true predictors for embryo euploidy. The best way to avoid misclassification of embryos due to univariate or subjective embryo scoring is to perform prospective, large-scale studies with several embryo variables, resulting in a scoring algorithm that is based on multivariate statistical analyses, in which all relevant variables ‘compete’. Such an algorithm thus includes only the factors that show independent impact on treatment outcome and it also gives the correct power balance between those variables. In the present study we included all treatments, also from repeated cycles, as this strategy may better reflect a couple’s

<table>
<thead>
<tr>
<th></th>
<th>linBL</th>
<th>abs(BL-4)</th>
<th>NU</th>
<th>FR</th>
<th>EQ</th>
<th>SY</th>
</tr>
</thead>
<tbody>
<tr>
<td>linBL (r)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abs(BL-4) (r)</td>
<td>-0.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td>&lt;0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NU (r)</td>
<td>0.02</td>
<td>-0.40</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td>0.050</td>
<td>&lt;0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR (r)</td>
<td>-0.03</td>
<td>0.16</td>
<td>-0.23</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td>0.012</td>
<td>&lt;0.000</td>
<td></td>
<td>&lt;0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ (r)</td>
<td>0.07</td>
<td>0.25</td>
<td>-0.35</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td>&lt;0.000</td>
<td>&lt;0.000</td>
<td></td>
<td>&lt;0.000</td>
<td>&lt;0.000</td>
<td></td>
</tr>
<tr>
<td>SY (r)</td>
<td>0.07</td>
<td>-0.02</td>
<td>-0.11</td>
<td>0.05</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>(P)</td>
<td>&lt;0.000</td>
<td>0.213</td>
<td>&lt;0.000</td>
<td>0.003</td>
<td>&lt;0.000</td>
<td></td>
</tr>
</tbody>
</table>

Owing to its non-linear relationship with live birth rate, Blastomere (BL) is represented by two variables: linBL, the total number of blastomeres; and abs(BL-4), the absolute figure for the deviation from the ideal four blastomeres. r, coefficient of correlation; n = 6252 embryos.
full fertility potential, compared with restricting the data analysis to the first treatment only (Broekmans et al., 2006; Holte et al., 2011a; Brodin et al., 2013). We applied the GEE model which statistically corrects for the correlation between repeated treatments within couples (Zeger et al., 1988). Refraining from correcting for such bias may alter c statistic levels, in some cases markedly (Racowsky et al., 2009).

Cleavage rate was the most powerful predictor for live birth. The ideal cleavage rate on Day 2 was four blastomeres. This finding corroborates several other studies (Holte et al., 2007; Machtinger and Racowsky, 2013). The slight reduction in implantation and LBRs associated with a more rapid cleavage rate (not as marked as for slow cleavage) confirms our own (Holte et al., 2007) and other previous findings (Cummins et al., 1986; Claman et al., 1987; Staessen et al., 1992; Giorgetti et al., 1995; Ziebe et al., 1997; Van Royen et al., 1999; Magli et al., 2001). This complex, non-linear relationship between treatment outcome and blastomere number is important to capture mathematically to give a high accuracy of the algorithm. The finding was identical to our previous finding constructing the IMC, and therefore the blastomere relationship with treatment outcome was again best described by a composite variable (linBL and abs(BL-4)) (Holte et al., 2007).

The proportion of mononucleated blastomeres had almost as strong predicting power for live birth as cleavage rate. The linear prognostic impact of the embryo nuclear status, with multinucleation being the least favourable predictor, was first shown in our previous paper (Holte et al., 2007). Multinuclearity is an established risk marker for embryo abnormality and decreased implantation potential (Tesarik et al., 1987; Van Royen et al., 1999, 2001, 2003). A recent study (Fauque et al., 2013), with 1629 treatments, demonstrated multinucleation as an independent negative variable after multivariate regression analysis, but could not confirm the proportion of mononucleated blastomeres as an independent variable, although the variable was significant in the univariate analysis.

This study shows that degree of fragmentation qualifies as an independent significant predictor of live birth in the multivariate model, as a marginally less strong marker than cleavage rate and on the same power level as nuclear status. As is illustrated in Fig. 1, the variable shows a similar linear relationship with LBR, as does nuclear status, with subtle but visible effects also at low degree fragmentation. This effect is also illustrated in the ranking tree for the rIMC score with slight fragmentation resulting in a somewhat lower rIMC score (9.1 versus 10 at ideal BL 4 and NU 3; Supplementary Fig. S1), corresponding to a reduction in observed and predicted LBRs of 3–4%.

The variation in blastomere size has been observed to have prognostic importance in earlier studies (Giorgetti et al., 1995; Ziebe et al., 1997; Antczak and Van Blaricom, 1999; Hardarson et al., 2001; Rienzi et al., 2005; Holte et al., 2007). The present study could not demonstrate blastomere size variation as an independent marker; as opposed to the results in our previous study (Holte et al., 2007), although it was highly significant in univariate analysis. It should however be stressed that the difference with our previous results was subtle, resulting in a switch between including degree of fragmentation instead of blastomere size variation. It is difficult to draw solid biological conclusions from these differences, but they may be related to differences in the two study protocols (SET versus DET, primary end-points LBRs versus implantation), and differences in size of study populations. Fragmentation showed a weaker association with blastomere number than did the variation in blastomere size (Table V), a factor of possible statistical importance when qualifying as an independent variable together with the strongest predictor blastomere number in the multivariate model. Thus, our findings should not discourage from scoring blastomere size variation, given the previous support for a biological significance of this variable per se, although it did not qualify in the final ranking model for predicting live birth. In fact, applying our previous IMC model (with blastomere size variation instead of degree of fragmentation in the model) to the present data resulted in discriminatory levels close to those attained with the new model (data not shown).

Symmetry of cleavage is the variable in this study with the least previous scientific support, and it was the only variable which showed no significant association with LBR also in univariate testing, confirming our previous results (Holte et al., 2007). Together the findings suggest this variable may be omitted when scoring embryos in the early cleavage stage.

The best predictive capacity (c statistic) of live birth in our cohort was 0.579, with the variables BL (linBL and abs(BL-4)), NU and FR. It should be stressed that this relatively low discrimination level resulted from only measures of embryo quality. Nevertheless, the c statistic we reached for the rIMC score only corrected for age is comparable to the discrimination levels attained in most full IVF prediction models (Leushuis et al., 2009; van Loendersloot et al., 2013). Other variables, such as measures of ovarian reserve or ovarian response to FSH/hMG (Holte et al., 2011b; Brodin et al., 2013; Huber et al., 2013), and a variable covering outcomes of previous IVF attempts would probably show independent effects above those in the present model and consequently increase the total discriminatory capacity (Leushuis et al., 2009; Holte et al., 2011a). However, the aim of this study was not to construct a full prediction model for counselling of patients. The interpretation of the present results should primarily focus on the rank order scheme for optimal selection within a cohort of embryos. Furthermore, there is a growing recognition that the c statistic, which plays a central role in evaluating diagnostic tests, has limitations in the evaluation of prediction models (Cook, 2007). Modest discriminative qualities may still result in clinically valuable calibrations at the group level (Cook, 2007; Leushuis et al., 2009; Te Velde et al., 2014). The present model did indeed show a strikingly good calibration on the group level over the whole range of predicted probabilities of pregnancy and live birth, also when subjected to a strictly conservative statistical test. Thus, the rIMC alone could correctly differentiate treatment prognosis in groups of patients in 10% stratas from low to high chance in this large population. Such good calibration has not been shown for other prediction models, even when other treatment variables were included and the ambition was to create a full prediction model (Te Velde et al., 2014). This finding underlines the great importance of an accurate embryo scoring protocol for ranking and selecting embryos to transfer, and suggests that such precise and standardized scoring should be the cornerstone of full prediction models.

The results must, however, be interpreted with caution. One limitation of the study involves the risk for overestimation when a prediction model is created and validated in the same data set. However, when tested with a specific statistical method, which controls for such bias (Harrell et al., 1996; Moons et al., 2004), we observed no decrease in statistical power. Nevertheless, the model should be validated in other independent populations (Te Velde et al., 2014).

It could also be seen as a limitation of this study that we studied only Day 2 transfers. However, short time of culture may be associated with lower risks for the offspring compared with prolonged culture, as recently suggested by three large national follow-up studies (Kallen et al., 2010; Kaira et al., 2012; Dar et al., 2013). Indeed, a major argument
for blastocyst culture is the possibility to perform more accurate embryo selection at that stage. Given the excellent calibration for the present model, it should be possible to switch to early cleavage transfers with the use of the rIMC. Other limitations of the present study are those inherent for real-time visual scoring including risks of inter-observer variation and the hazards of fixed time-point scoring procedures in a dynamic process. Time-lapse recordings have so far resulted in important results for deselection of aberrant cleaving embryos (Herrero et al., 2013). Some results indicate that time-lapse recording may also prove to be powerful for positive selection between embryos of otherwise similar morphology and cleavage stage (Aparicio et al., 2013; Conaghan et al., 2013; Herrero et al., 2013; Montag et al., 2013). It would be of great interest to combine the present scoring model with the time-lapse technique to further optimize embryo selection.

In conclusion, the findings of this large prospective study on visual embryo scoring variables in SET show that blastomere number (in a non-linear way), the proportion of mononucleated (including information on multinucleation) blastomeres and the degree of fragmentation (both linearly) have independent prognostic power to predict live birth. Together, these variables are incorporated in an algorithm, which is presented as the rIMC score. It is suggested that applying the ranking tree, as presented here, gives an easy and structured clinical guide to the ranking and selection of embryo/s most suitable for transfer in the early cleavage stage.

**Supplementary data**

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

**Authors’ roles**

A.R.: participated in the study design, the interpretation of the data and wrote the main parts of the manuscript in cooperation with the other authors. I have had full access to all the data in the study and had final responsibility for the decision to submit for publication. L.B.: supervised the statistical evaluation and advised in interpretation of the data. He has also been involved in the design of the study. T.B.: participated in collecting the data, the design of the study and has taken part in supervising the development of the manuscript. M.O.: interpretation of data; supervised the development of the article and advised in the writing of the manuscript. K.M.: participated in the study design, performed active scoring of the embryos and has taken part in revision of the article and final approval. N.H.: performed statistical analysis, participated in study design and interpretation of data. J.H.: conception of the study and the study design, interpretation of the data and revised the manuscript for final approval.

**Funding**

Carl von Linne Clinic, Uppsala and the Department of Women’s and Children’s Health and the Family Planning Fund in Uppsala, Uppsala University, Uppsala, Sweden financed this study.

**Conflict of interest**

None declared.

**References**


Hardarson T, Hanson C, Spjøren A, Lundin K. Human embryos with unevenly sized blastomeres have lower pregnancy and implantation rates.


Kaira SK, Ratcliffe SJ, Barnhart KT, Coutifaris C. Extended embryo culture and an increased risk of preterm delivery. *Obstet Gynecol* 2012; 120:69–75.


