Timing of human preimplantation embryonic development is confounded by embryo origin

K. Kirkegaard[1,2,*], L. Sundvall[2,3], M. Erlandsen[4], J.J. Hindkjær[3], U.B. Knudsen[2,3], and H.J. Ingerslev[2,3]

[1]Department of Clinical Biochemistry, Aarhus University Hospital, Nørrebrogade 44, DK-8000 Aarhus C, Denmark [2]Clinical institute, Health, Aarhus University, Palle Juul Jensens Boulevard, 8200 Aarhus N, Denmark [3]Centre for Preimplantation Genetic Diagnosis/The Fertility Clinic, Aarhus University Hospital, Palle Juul Jensens Boulevard, 8200 Aarhus N, Denmark [4]Department of Public Health, Section for Biostatistics, Aarhus University, Aarhus, Denmark

*Correspondence address. E-mail: kirstine.kirkegaard@clin.au.dk

STUDY QUESTION: To what extent do patient- and treatment-related factors explain the variation in morphokinetic parameters proposed as embryo viability markers?

SUMMARY ANSWER: Up to 31% of the observed variation in timing of embryo development can be explained by embryo origin, but no single factor elicits a systematic influence.

WHAT IS KNOWN ALREADY: Several studies report that culture conditions, patient characteristics and treatment influence timing of embryo development, which have promoted the perception that each clinic must develop individual models. Most of the studies have, however, treated embryos from one patient as independent observations, and only very few studies that evaluate the influence from patient- and treatment-related factors on timing of development or time-lapse parameters as predictors of viability have controlled for confounding, which implies a high risk of overestimating the statistical significance of potential correlations.

STUDY DESIGN, SIZE, DURATION: Infertile patients were prospectively recruited to a cohort study at a hospital fertility clinic from February 2011 to May 2013. Patients aged <38 years without endometriosis were eligible if ≥8 oocytes were retrieved. Patients were included only once. All embryos were monitored for 6 days in a time-lapse incubator.

PARTICIPANTS/MATERIALS, SETTING, METHODS: A total of 1507 embryos from 243 patients were included. The influence of fertilization method, BMI, maternal age, FSH dose and number of previous cycles on timing of t2-t5, duration of the 2- and 3-cell stage, and development of a blastocoel (tEB) and full blastocoel (tFB) was tested in multivariate, multilevel linear regression analysis. Predictive parameters for live birth were tested in a logistic regression analysis for 223 single transferred blastocysts, where time-lapse parameters were investigated along with patient and embryo characteristics.

MAIN RESULTS AND THE ROLE OF CHANCE: Moderate intra-class correlation coefficients (0.16–0.31) were observed for all parameters except duration of the 3-cell stage, which demonstrates that embryos from one patient elicit clustering at a patient level. No single patient- and treatment-related factor was found to systematically influence the timing from cleavage to blastocyst stage, which indicates that no individual patient-related factor can be identified that separately explains the clustering throughout the entire developmental stages. The blastocyst parameters were more affected by patient-related factors than cleavage stage parameters, as tEB occurred significantly later with older age (0.29 h/year (95% confidence interval: CI 0.03; 0.56)), while both tEB and tFB occurred significantly later with increasing dose of FSH (tEB: 0.12 h/100 IU FSH (95% CI 0.01;0.24); tFB 0.14 h/100 IU FSH (95% CI 0.03;0.27)) and with more previous attempts (tEB: 1.2 h/attempt (95% CI 0.01;2.5); tFB 1.4 h/attempt (0.10;2.7)). Fertilization method affected timing of the first division, with ICSI embryos cleaving significantly faster than IVF embryos (−3.6% (95% CI −6.4; −0.77)), whereas no difference was found in the subsequent divisions. The univariable regression analysis identified female age, cumulative FSH dose, degree of blastocyst expansion, score of the inner cell mass and timing of full blastocyst formation as predictors of live birth. The timing of full blastocyst formation (tFB) did not remain significant when adjusting for age, number of previous cycles and cumulative FSH dose, which were the parameters shown to influence tFB in the mixed regression model.

LIMITATIONS, REASONS FOR CAUTION: Only good prognosis patients were enrolled, so these results may not be generalized to all infertile women. Not all patient-related factors were investigated.

WIDER IMPLICATIONS OF THE FINDINGS: Our findings underline the importance of treating embryos as dependent observations and suggest a high risk of patient-based confounding in retrospective studies. The impact of confounders and the embryo origin needs to be addressed in order to apply appropriate statistical models in observational studies. Furthermore, this observation emphasizes the need for RCTs for evaluating use of time-lapse parameters for embryo selection.

STUDY FUNDING/COMPETING INTERESTS: Funding for the cohort study was provided by the Lippert Foundation, the Toyota Foundation, the Aase og Einar Danielsen foundation and NordicInfu Care research grant. Research at the Fertility Clinic, Aarhus University Hospital is supported by an unrestricted grant from MSD and Ferring. K.K. is funded by a grant from the Danish Council for Independent Research Medical Sciences. The authors declare no competing interest.

Key words: embryo development / human / time-lapse / embryo transfer / pregnancy / confounding

Introduction

The use of time-lapse incubators designed to support and monitor the development of IVF embryos has led to a rapid introduction of time-lapse monitoring (TLM) in clinical practice. The implementation of TLM for routine use in the IVF laboratory is based on the assumption that uninterrupted culture and improved embryo selection will lead to increased pregnancy rates. Studies that report a relationship between timing of development and embryo viability are, however, mostly observational and, as such, at risk of confounding. A recent RCT (Rubio *et al.*, 2014) followed up on the encouraging results from several retrospective studies evaluating a hierarchical selection model (Meseguer *et al.*, 2011, 2012). The trial confirmed the positive findings by reporting an odds ratio (OR) of 1.23 (1.06–1.43) in favour of TLM compared with standard incubation and selection for the primary outcome, namely ongoing pregnancy. However, the patients received two interventions as they were not only randomized to two different selection procedures, but also to different culture systems, which precludes a separate evaluation of the effect of time-lapse selection *per se*. The above mentioned hierarchical selection model is by far the most well investigated and only a few other clinically applicable models have been proposed (Campbell *et al.*, 2013a; Conaghan *et al.*, 2013; VerMilyea *et al.*, 2014). None of these models have been tested in randomized trials. The evidence to support the assumption that TLM *per se* improves embryo selection is therefore still weak (Kaser and Racowsky, 2014; Armstrong *et al.*, 2015a; Kirkegaard *et al.*, 2015). Furthermore, concerns have been raised as to whether the tested models are transferable to different clinical settings (Kirkegaard *et al.*, 2014a; Freour *et al.*, 2015), which has promoted the view that each team using time-lapse technology should build a centre-specific prediction model based on its own data and transfer policy (Freour *et al.*, 2015). This recommendation is based on a number of publications suggesting an influence on timing of a variety of patient- and treatment-related factors (Ciray *et al.*, 2012; Munoz *et al.*, 2012, 2013; Bellver *et al.*, 2013; Cruz *et al.*, 2013; Freour *et al.*, 2013; Kirkegaard *et al.*, 2013a). As is the case for most studies that correlate time-lapse parameters to blastocyst development or aneuploidy (Cruz *et al.*, 2012; Dal Canto *et al.*, 2012; Campbell *et al.*, 2013a; Conaghan *et al.*, 2013; Basile *et al.*, 2014), only few of the publications have controlled for confounding and most of the studies have treated embryos from each patient as independent observations. As this approach implies a risk of overestimating potential correlations, it is highly relevant to address the potential pitfalls of confounding and clustering when evaluating time-lapse data. The aim of this study was therefore to establish to what degree different patient-related factors influence the timing of embryo development, and to investigate whether variation in embryo development is patient dependent by performing a multi-level, multi-variable analysis of development of IVF embryos to the blastocyst stage.

Materials and Methods

Study design and participants

Embryos from infertile women undergoing IVF or ICSI treatment at the Fertility Clinic, Aarhus University Hospital were recruited prospectively to a cohort study from February 2011 to May 2013. During this period, patients regarded as good prognosis patients and undergoing IVF/ICSI treatment were offered participation if the woman was aged <38 years and had no endometriosis. Absence of endometriosis was based on a combination of clinical information (no dysmenorrhoea) and ultrasound (no endometrioma). The embryos were included when patients had given signed informed consent if ≥8 oocytes were retrieved. Eligible patients could contribute to the study with one treatment cycle only. Data related to patient characteristics were obtained for the current treatment cycle. The present paper presents a multivariate analysis of development in the embryos from 243 patients. Previous publications on the same cohort or subgroups of the cohort have reported on the correlation between early time-lapse parameters and embryo development (Kirkegaard *et al.*, 2013b), correlation between metabolism and clinical outcome (Kirkegaard *et al.*, 2014b) and development of embryos from women with polycystic ovary syndrome (Sundvall *et al.*, 2015).

Ethical approval

Written informed consent was obtained from all participants before inclusion. The Central Denmark Region Committees on Biomedical Research Ethics and the Danish Data Protection Agency approved the study. The study was registered at ClinicalTrial.gov with accession number NCT01139268 and the prolongation with accession number NCT01953146.

IVF/ICSI, embryo culture and collection of media samples

Ovarian stimulation and oocyte retrieval were performed according to standard procedures as previously described (Kirkegaard *et al.*, 2012b). Following retrieval, oocytes were placed in Cook fertilization medium

(COOK®, Australia) and fertilized with conventional ICSI or IVF procedures. Indications for ICSI were male infertility or three previously failed IVF fertilization attempts. ICSI fertilized embryos were placed in individual wells (EmbryoSlide, Fertilitech, Aarhus, Denmark) immediately after injection and cultured in a tri-gas time-lapse incubator (EmbryoScope, Fertilitech, Aarhus, Denmark). IVF embryos were cultured for ∼18 h in a conventional incubator (Galaxy R, RS Biotech, CM Scientific, West Lothian, UK) under oil at $37°C$, 20% O_2 and 6% CO_2 before transfer to the EmbryoScope. In the EmbryoScope, embryos were cultured at $37°C$, 5% O_2 and 6% CO_2 in 25 µl individual wells containing droplets of Sydney IVF Cleavage Medium (COOK®, Australia) under oil, with a change to Sydney IVF Blastocyst Medium (COOK®, Australia) 68 h (Day 3) after fertilization. On Day 5, a trophectoderm (TE) biopsy was obtained for research purposes from 23 of the transferred embryos. In each cycle the single embryo with the highest morphological grade was selected for transfer in the morning of Day 6 after morphological evaluation in an inverted microscope at $200×$ magnification and grading blastocysts according to the Gardner criteria; in brief based on expansion of the blastocoel cavity (1–6), number and cohesiveness of the inner cell mass (ICM) and TE (A-C). No time-lapse parameters were used in the selection process. All transfers were performed on Day 6 according to the protocol, motivated by the requirement for regeneration in case of a biopsy (Kokkali et al., 2005). Biochemical pregnancy was confirmed by serum β-hCG measurement 16 days after oocyte retrieval. Clinical pregnancy rate was confirmed by ultrasound as presence of fetal heart activity 8 weeks after embryo transfer. Live birth was recorded as the birth of a child.

Time-lapse imaging and assessment

Images were recorded automatically in seven focal planes every 20 min (15 µm intervals, $1280 × 1024$ pixels, 3 pixels per µm, monochrome, 8-bit < 0.5 s per image, using single 1W red LED). A time-point was automatically assigned to each image reported as hours after time zero (t0). For ICSI embryos t0 was defined as the time of injecting the sperm into the oocyte. For IVF embryos t0 was defined as the time of adding the sperm to the dish. Manual annotation of time-lapse images was performed at an external workstation (Embryo Viewer™). The time-lapse parameters were annotated according to definitions previously described (Kirkegaard et al., 2012a, 2013b; Sundvall et al., 2013). The time-lapse parameters included in the present analysis were time of first division (t2), time of division to three cells (t3), time of division to five cells (t5), duration of the 2-cell stage (t3-t2), duration of the 3-cell stage (t4-t3), start of formation of a blastocoel (tEB), and time of full blastocoel formation (tFB). If evaluation of specific events was not possible due to unfocused imaging, oil drops or technical problems such as no recording, these data points were treated as missing data. Two observers performed the time-lapse annotations. Inter-observer variability of the analysis has been evaluated in another study (Sundvall et al., 2013) where we reported average value of intra-class correlation coefficients (ICC) of 0.8 and the median value ICC of 0.9. Durations of cell cycles and cell stages were subsequently calculated as the interval between two time-points.

Statistical analysis

The time-lapse data were analysed using a multilevel mixed-effects linear regression model to control for the multilevel random and systemic variation. The model was chosen to account for the assumption that embryos from each patient are biologically more similar than when comparing embryos between patients (intra- versus inter-individual variation, respectively), i.e. the observations (embryos) are grouped into clusters according to patient origin. Mixed models can account for the correlation among observations in the same cluster, and give an estimate of this correlation. In our model the systematic effects are considered to vary according to the patient: ICCs were calculated in order to quantify the clustering effect. The ICC estimates the proportion (0–1) of the total variance of the timing parameter that is accounted for by the patient origin and therefore indicates whether the observations cluster (and therefore cannot be analysed as independent observations) and if the patient origin is expected to influence the timing (i.e. is a confounder). The higher the ICC, the less unique information each additional observation provides. In general, a double digit ICC justifies a multilevel analysis.

The explanatory variables included in the model for the analysis including all embryos were: female age, method of fertilization, BMI, cumulative FSH dose and number of previous cycles. The end-points evaluated (t2-t5, tEB, tFB, and duration of the 2- and 3-cell stage) and the explanatory variables were chosen based on the literature. Our criteria for selecting the explanatory variables were that the variables should either be known confounders (such as age and number of previous cycles) or have been suggested as having an impact on timing (BMI, fertilization method, total FSH). The outcome variables were chosen based on the following considerations: t2 was chosen as early cleavage (which to some extent is reflected by t2), which has traditionally been considered a relevant predictor. t3 and t5 were chosen as they are important parameters in the hierarchical model (t5) or the adjusted hierarchical model (t3 and t5) (Meseguer et al., 2011; Basile et al., 2015). t3-t2 and t4-t3 were chosen as they are included in the hierarchical model and constitute the basis for the second most investigated model (Conaghan et al., 2013; VerMilyea et al., 2014). tEB/tFB was chosen to include blastocyst parameters, and form the basis of a published aneuploidy prediction model (Campbell et al., 2013a,b).

The model assumptions of linearity, identical distributions and normality of the random deviations were checked by probability plots and by plotting residuals against the predicted values. If the model did not fulfil the assumptions continuous data were log transformed, and estimates are reported as predicted percentage difference in timing of the event per unit of the variable evaluated (age, IVF/ICSI, BMI, FSH, previous cycles) with 95% confidence intervals (CI). The estimates thus report what percentage the timing of the investigated event (t2, t3, etc.) is expected to occur faster or slower with a change in 1 unit of the variable, for example 1 year (age), 100 IU FSH, ICSI compared with IVF. Estimates for non-transformed continuous data (tEB, tFB) are reported as predicted difference in hours per unit of the variable (age, IVF/ICSI, BMI, FSH, previous cycles) with 95% CI. The estimates thus report how many hours the event (tEB, tFB) is predicted to occur earlier or later with a change in 1 unit of the variable, for example 1 year (age), 100 IU FSH, ICSI rather than IVF. For direct cleavage (binary outcome), estimates are reported as OR.

As a supplementary analysis, the ability of time-lapse parameters to predict pregnancy was tested by obtaining ORs with the use of a logistic regression model evaluating the chosen time-lapse parameters as predictors of clinical pregnancy.

Baseline data were tested for the assumption of normality by histograms and QQ plots (plotting quantiles of the first data set against quantiles of the second data set). If continuous data did not fulfil the assumption of normality, a Wilcoxon Rank Sum Test analysis was performed and estimates are reported as medians and range. Normally distributed continuous data were analysed using student's t-test, and estimates are reported as mean ± SD and range. Categorical data were analysed with Fishers exact test.

All statistical analyses were performed in the statistical package STATA for Mac, version 14.0 (StataCorp, USA). Two-sided P-values <0.05 were considered significant.

Results

In total, embryos from 243 patients were included in the study. Baseline data for these 243 patients are listed in Table I. In order to evaluate the development of competent embryos, we excluded immature oocytes,

abnormally fertilized oocytes and embryos that had not reached the 4-cell stage within an arbitrarily defined limit of 60 h, which left 1507 embryos for the primary multivariate, multilevel linear regression analysis of development.

Results from the multilevel mixed-effects linear regression analysis are found in Table II. Moderate ICCs (0.16–0.31) were observed for all parameters except duration of the 3-cell stage.

In general, no single patient- and treatment-factor was found to elicit a systematic influence on the overall timing from the cleavage until the blastocyst stage. However, the analysis suggested an influence on some of the timings. In particular, the blastocyst parameters appeared to be more affected by the patient-related factors than cleavage stage parameters, as tEB occurred significantly later with older age, while both tEB and tFB occurred significantly later with increasing dose of FSH and with more previous IVF/ICSI attempts (Table II). Fertilization method affected only timing of the first division, with ICSI embryos cleaving significantly faster than IVF embryos, whereas no difference was found in the subsequent divisions (Table II).

For 20 patients, embryo transfer was cancelled and 223 single trans-ferred blastocysts were therefore included in a supplementary pregnancy outcome analysis. Patient and cycle characteristics for the pregnant and the non-pregnant groups are listed in Supplementary Table SI. The patients in the pregnant group were younger, had more embryos cryo-preserved and had an embryo of better quality transferred than patients in the non-pregnant group. Figure 1 displays the distribution of selected time-lapse parameters, which are very similar for embryos resulting in live birth and no live birth. The univariable regression analysis identified

Table I Baseline characteristics for the patients and cycles included in the time-lapse analysis of the embryo cohort.

Number of patients	243
Number of cycles	243
Number of previous cycles	0 (0;3)
Maternal age (years)	31 (20;37)
Maternal BMI (kg/m^2)	22.6 (16.2;38.9)
Indication for cause of infertility, n (%)	
Male	129 (53%)
Female	37 (15%)
Unexplained	77 (32%)
Cumulative FSH dose (IU)	1650 (200;4875)
Aspirated oocytes	12 (8;34)
Fertilization method	
Standard IVF	92
ICSI	151
Number of embryos with 2 pronuclei	6 (0;23)
Number of cycles with embryo transfer	223
Biochemical pregnancies	91/223 (40,8%)
Fetal heart beat	67/223 (30,0%)
Live birth	64/223 (28,7%)
Embryos cryopreserved	2 (0;10)

Continuous data are presented as medians and range.

female age, cumulative FSH dose, degree of blastocyst expansion, score of the ICM and timing of full blastocyst formation as predictors of live birth (Supplementary Table SI). Timing of full blastocyst formation did not remain significant when adjusting for age, number of previous cycles and cumulative FSH dose, which were the parameters shown to influence tFB in the mixed regression model.

Discussion

To our knowledge, this study is the first to perform an extensive analysis of the effect of embryo origin on preimplantation embryo development until Day 6. This study serves to illustrate two important points with rather crucial implications for the design and interpretation of TLM studies. Firstly, embryos from one patient elicit clustering, which means that a large part of the variation observed between embryos can be explained by differences between patients. Secondly, as the indi-vidual origin of the embryos influences the timing of the development, embryo origin must be considered a potential source of confounding. With a few exceptions no individual factor can be identified that separ-ately explains the variation throughout the entire development from cleavage until blastocyst stage.

The first statement above regarding clustering is substantiated by the moderate ICCs that were observed for almost all the parameters evaluated. The ICC is particularly useful in linear mixed models. These models account for the correlations among observations in the same cluster, and give an estimate on how much of the overall variation is explained simply by clustering. In this model we tested whether timing of selected parameters cluster on a patient level. The ICC thus gives an es-timate of how much of the variation in timing can be explained by the patient origin. We found that between 16 and 31% of the observed vari-ation in timing was explained by the grouping variable, which in our model was the patient from which the embryos originated. Our findings thus demonstrate that embryos originating from one patient are more similar in their developmental timing compared with embryos from other patients. As a consequence cohorts of embryos from individual patients cannot be treated as independent observations. The clustering introduces a design effect, which arises since additional observations do not provide unique information. The design effect decreases the power of the study if clustering is present. The degree of the design effect depends on the ICC and the average cluster size (n) (Design effect = $1 + (n - 1)*$ICC) (Kirkwood and Sterne, 2003). Accordingly, with ICC in the range of 0.15–0.30, the design effect can easily reduce power sub-stantially. Consequently, TLM studies that include more than one embryo from each patient—which are common in studies of blastocyst prediction, prediction of aneuploidy and evaluation of different patient and treatment-related factors—should be based on statistical models that account for the grouping of embryos. Accordingly, the use of any test that assumes embryo observations to be independent, such as a Wilcoxon Rank-sum test and the equivalent parametric Student t-test, carries a high risk of overestimat-ing potential correlations. To illustrate the implications of violating the assumptions of independency we invite you to consider the following example: The graphical presentation of IVF versus ICSI embryos suggests that the timings in the two groups are nearly identical (Supplementary Fig. S1). If a standard non-parametrical test (Wilcoxon rank-sum test) is used to compare timing of development between IVF and ICSI fertilized embryos, the test produces highly significant differences for almost all eval-uated parameters, in line with other publications using similar methods

Table II Results of the mixed effect linear regression model for all human embryos.

	Age (years)	Fertilization method (ICSI compared with IVF)	BMI (kg/m^2)	Total FSH (100 IU)	Number of previous cycles	ICC
t2 (%) $n = 1324$	0.21 ($-0.23;0.65$)	-3.6^* ($-6.4;\ -0.77$)	0.42 ($-6.6;\ 8.0$)	0.18 ($-0.02;0.37$)	1.5 ($-0.5;3.6$)	0.31 (0.25;0.37)
t3 (%) $n = 1366$	0.26 ($-0.18;0.71$)	-2.4 ($-5.1;0.50$)	5.1 ($-22.0;13.0$)	0.10 ($-0.10;0.29$)	0.83 ($-1.2;2.9$)	0.25 (0.20;0.32)
t4 (%) $n = 1361$	0.16 ($-0.27\ ;0.60$)	-2.1 ($-4.8;0.74$)	-1.1 ($-7.9;\ 6.2$)	0.10 ($-0.10;0.29$)	2.0 (0.00;4.1)	0.21 (0.16;0.27)
t5 (%) $n = 1298$	0.36 ($-0.10;0.83$)	-1.5 ($-4.4;1.5$)	0.83 ($-6.5;8.7$)	0.10 ($-0.12;0.29$)	0.83 ($-1.2;2.9$)	0.25 (0.20;0.31)
t3-t2[#] (%) $n = 1155$	0.32 ($-0.13;0.77$)	-2.4 ($-5.3;0.52$)	-2.9 ($-9.9;4.5$)	-0.05 ($-0.25;0.15$)	1.5 ($-0.58;3.7$)	0.21 (0.16;0.28)
t4-t3 (%) $n = 1360$	-0.65 ($-2.2;\ 0.9$)	1.8 ($-8.3;12.9$)	-32^* ($-48;-12$)	0.10 ($-0.62;0.83$)	11^* (3.3;20.1)	0.03 (0.01;0.09)
tEB (hours) $n = 1061$	0.29^* (0.03; 0.56)	-1.37 ($-3.1;0.34$)	0.36 ($-3.9;4.7$)	0.12^* (0.01;0.24)	1.2^* (0.01;2.5)	0.19 (0.13;0.27)
tFB (hours) $n = 1020$	0.21 ($-0.07;0.49$)	-1.76 ($-3.6;.04$)	-0.31 ($-4.8;4.2$)	0.14^* (0.03;0.27)	1.4^* (0.10;2.7)	0.19 (0.13;0.27)
Direct cleavage (OR) (t3-t2 < 5 h) $n = 1387$	0.99 (0.92;1.06)	0.88 (0.57;1.4)	0.13^* (0.04;0.45)	1.0 (0.98;1.0)	1.4 (1.0;1,8)	0.16 (0.08;0.29)

Estimates from analysis of ln transformed data (all analyses except tEB and tFB) are reported as % difference in timing per unit variable (age, ICSI compared with IVF, BMI, FSH, previous cycles) (95% confidence interval (CI)). Estimates for non-transformed continuous data (tEB, tFB) are reported as predicted difference in hours per unit variable (age, ICSI compared with IVF, BMI, FSH, previous cycles) (95% CI). For direct cleavage (binary outcome), estimates are reported as odds ratio (OR). ICC = Intra Class Correlation.
t2: time of division to two cells. t3: time of division to three cells. t5: time of division to five cells. t3-t2: duration of the 2-cell stage. t4-t3: duration of the 3-cell stage. tEB: start of formation of a blastocoel. tFB: time of full blastocoel formation.
$^*P < 0.05$ (mixed effect linear regression).
[#]Only embryos not displaying direct cleavage.

(*t*-test, Wilcoxon rank-sum test) (Cruz *et al.*, 2013; Bodri *et al.*, 2015). This clear discrepancy between the graphical presentation of data and the test is verified when data are evaluated in the more appropriate multivariate and multilevel model that accounts for dependent observations and confounding (Table II, Supplementary Fig. S1), where we find that only the first division is significantly influenced by fertilization method. From this example we can infer that previous observational studies—including studies from our own group—which have not accounted for clustering are at high risk of having overestimated the reported influences of the individual external factors.

The moderate ICCs indicate that embryo development is correlated to patient origin. This forms the basis for our second statement, that patient origin must be considered a confounder in observational studies. While clustering will affect only studies where cohorts of embryos are evaluated and not studies where only a single embryo is evaluated, the risk of confounding will affect all retrospective time-lapse studies, including studies evaluating single transferred embryos. The general risk of confounding naturally applies to all observational studies, including studies of the predictive value of standard morphology, and is mostly accounted for by using multivariate models. If the patient origin is not accounted for in observational studies, it cannot be concluded whether the observed differences in timing between implanted and non-implanted embryos in reality reflect differences between different groups of patients with different prognosis. This finding illustrates the need for a patient-based approach and the importance of validating observational studies with RCTs. As an example, our univariable analysis suggests that timing of full blastocyst formation (tFB) is correlated with pregnancy (Supplementary Table SI). As a consequence, we could suggest a prediction model, where tFB is used as a selection parameter. Our multivariate regression analysis also demonstrates, however, that the exact same parameter (tFB) is correlated with age (Table II), a

parameter that is known to correlate strongly with pregnancy (van Loendersloot *et al.*, 2010). When adjusting for age in the multivariable logistic regression analysis, tFB formation was no longer an independent predictor of pregnancy. This serves to illustrate that in case of such obvious confounding, different timings between embryos that implant and embryos that do not may not reflect an actual biological association between timing of embryo development and clinical outcome, but rather that different patients, in this example of different age, will have different chances for pregnancy. This can to some extent be corrected for in a multivariate analysis, if the confounders are known. However, as our analysis does not unequivocally answer which factors to control for, in particular for the cleavage stage parameters, the ultimate way to circumvent this challenge is to validate proposed time-lapse models by performing a randomization of patients, thereby securing an equal distribution of known, as well as unknown, confounders. The trial design would randomize women to either time-lapse incubation and selection, or time-lapse incubation with conventional assessment of morphological parameters (Armstrong *et al.*, 2015b).

Our cohort consisted of embryos from young patients (<38 years) with many embryos and no endometriosis, which we in general consider good prognosis patients. Although we were unable to identify single factors with an independent influence on timing, we found a clear grouping effect at the patient level. Our patients constitute a specific subgroup of selected patients. While the findings with regard to the supplementary analysis of the outcome might not be transferrable to a broader population, we would expect that the clustering and confounding effects would be even more pronounced in a more heterogeneous patient population. It is plausible that some of the factors, such as age, would have demonstrated a significant influence on timing in a population with a wider age span. This would, in our opinion, suggest that observational studies with a more heterogeneous population than the present, for example
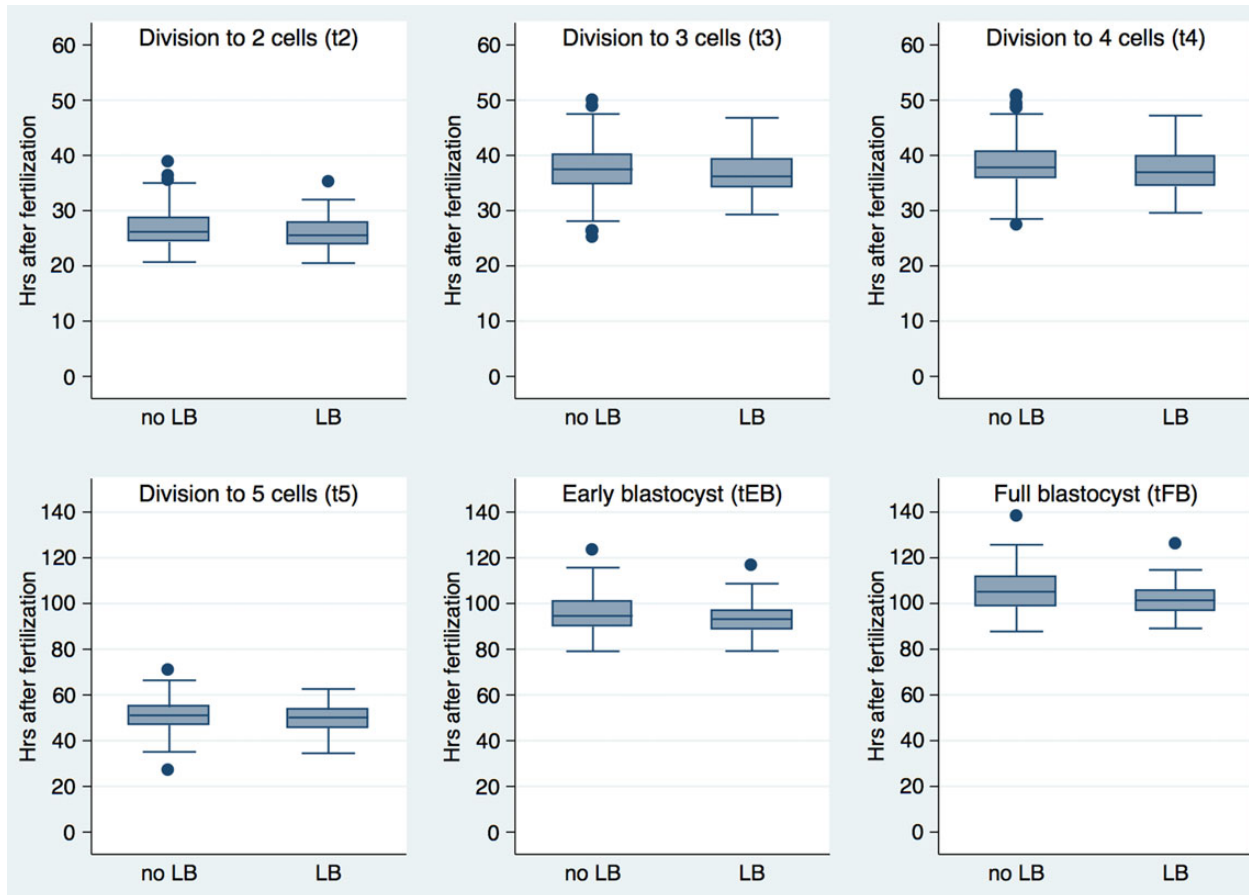
Figure 1 Time points of selected embryonic stages for human embryos resulting in live birth and no live birth. The middle band inside the box represents the median value, and the upper and lower limits of the box represent the upper and lower quartiles, respectively. Whiskers display the upper and lower values within 1.5 times the upper and lower quartiles. Outliers are displayed as dots. LB, live birth.

a wide age span or a large contribution from donors with a more favourable prognosis, would be even more prone to confounding, as a large part of the variation in embryo development is explained by patient- and treatment-related factors. This view is supported by a study (Bellver et al., 2013) where embryos derived from fertile donors divided significantly faster than embryos from infertile patients. Notably, the donors in this study were significantly younger, and had more mature oocytes and more embryos cryopreserved than the infertile population. This finding does, in our opinion, clearly illustrate the need to take the question of embryo origin seriously, and offers a possible explanation for why models that use time-intervals obtained from donors may be difficult to transfer to other patient populations.

In general, no single patient- and treatment-factor was found to elicit a systematic influence on the overall timing from the cleavage until the blastocyst stage, which complicates any controlling for confounding on timing and potentially the development of clinic-specific models. The analysis did, however, suggest an influence on some of the timings, in particular on the blastocyst development, which was influenced by age, number of previous cycles and cumulative FSH dose—factors which are mutually interrelated and all reflect ovarian function. Timing of blastocyst development has been proposed as a predictor of

aneuploidy (Campbell et al., 2013a,b), yet it was later questioned whether the observed differences in timing were the result of confounding by age (Ottolini et al., 2014). A later study found no correlation between neither cleavage stage nor blastocyst time-lapse parameters and aneuploidy using logistic mixed-effects models adjusted for age (Rienzi et al., 2015). Our study confirms that age is a significant confounder for blastocyst formation along with cumulative FSH dose and number of previous cycles as confounders. Previous studies have reported that IVF embryos display a systematic delay in development compared with ICSI embryos, that persists during the cleavage stages (Cruz et al., 2013; Bodri et al., 2015), whereas other studies have reported the delay to affect only during the first divisions (Dal Canto et al., 2012). As the difference at the cleavage stage disappeared when the timings were normalized to pronuclear fading rather than time of fertilization (Cruz et al., 2013; Bodri et al., 2015), the delays were hypothesized to arise from a later starting point for the IVF embryos, caused by the circumvention of normal sperm penetration in the ICSI procedure. As argued in the above section, the data were however analysed using t-tests, Wilcoxon Rank sum tests and chi-squared test, which do not account for clustering and confounding, and therefore carry a high risk of overestimating the correlation

observed. Our analysis indicates that the difference in timing between IVF and ICSI embryos is significant only for the first cleavage. We consider the most likely explanation to be the large variation in timing at the later stages, which makes the relatively small contribution from fertilization method disappear.

Although the aim of this study was not to evaluate time-lapse parameters as predictors of pregnancy, we conducted a supplementary logistic regression analysis on the transferred embryos in order to identify potential time-lapse predictors of live birth and to illustrate the importance of confounding factors. As illustrated graphically in Fig. 1, our logistic regression analysis revealed no difference in timing between live birth and non-implanted embryos for the cleavage stage parameters, which is in line with previous findings from a subgroup from the same cohort (Kirkegaard *et al.*, 2013b). Our results indicate that timing of blastocyst parameters may occur earlier in implanting embryos, as both blastocyst expansion and tFB were identified as predictors of live birth in our logistic regression analysis. The lack of difference in timing at the early cleavage stages in contrast to difference in timing after t5 is in concordance with the findings from a recent study identifying parameters for formation of good quality blastocysts applying generalized estimating equations, which adjust for correlation between observations (Storr *et al.*, 2015). Adjusting for age, Storr *et al.* (2015) found that tEB was the most significant predictor for a top-quality blastocyst. In contrast, our findings did not remain significant when adjusting for age, FSH dose and number of previous cycles. The external validity of the implantation analysis may be affected by the elective day 6 transfer policy, as blastocyst transfers in most clinics are performed on Day 5. Furthermore, a small subset of embryos was biopsied, which might potentially introduce an effect modification in the secondary outcome analysis. Regardless, the primary multilevel mixed-effects linear regression analysis addressing the impact of different factors until the full blastocyst stage is unaffected by the biopsy procedure, as the biopsy and laser-assisted opening of the zona pellucida were performed after the full blastocyst stage.

In conclusion, our study presents a detailed analysis of development to the blastocyst stage in a cohort of embryos from selected infertile women. Our analysis suggests that a high degree of the observed variation in development is patient dependent, and that the effect is most likely caused by a combination of several factors since no single factor elicits a systematic influence in our cohort. In our opinion TLM will prove useful for the IVF laboratory, as there are several potential benefits compared with standard morphological scoring such as improved laboratory workflow and culture conditions. While TLM may also improve our understanding of embryo development and selection, our findings underline the importance of treating embryos as dependent observations in studies where a variable number of embryos from each patient is analysed and suggest a risk of a high degree of confounding in all retrospective time-lapse studies, which emphasizes the need for RCTs when evaluating time-lapse parameters for embryo selection. Our aim is not to suggest that time-lapse selection will not prove useful, but merely to underline the need for a cautious interpretation of results and the use of appropriate statistical tools in all observational studies.

Supplementary data

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

Authors' roles

K.K., J.J.H., U.B.K. and H.J.I. designed the clinical study. K.K. and M.E. designed the statistical analysis. K.K. drafted the manuscript. L.S. and K.K. did the data acquisition and the time-lapse annotations and K.K. analysed and interpreted the data. K.K., H.J.I., L.S., M.E. and U.B.K. performed critical revision of the manuscript. All authors have given their final approval of present version to be published.

Conflict of interest

None declared.

References

Armstrong S, Arroll N, Cree LM, Jordan V, Farquhar C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev* 2015a;**2**:Cd011320.

Armstrong S, Vail A, Mastenbroek S, Jordan V, Farquhar C. Time-lapse in the IVF-lab: how should we assess potential benefit? *Hum Reprod* 2015b;**30**:3–8.

Basile N, Nogales Mdel C, Bronet F, Florensa M, Riqueiros M, Rodrigo L, Garcia-Velasco J, Meseguer M. Increasing the probability of selecting chromosomally normal embryos by time-lapse morphokinetics analysis. *Fertil Steril* 2014;**101**:699–704.

Basile N, Vime P, Florensa M, Aparicio Ruiz B, Garcia Velasco JA, Remohi J, Meseguer M. The use of morphokinetics as a predictor of implantation: a multicentric study to define and validate an algorithm for embryo selection. *Hum Reprod* 2015;**30**:276–283.

Bellver J, Mifsud A, Grau N, Privitera L, Meseguer M. Similar morphokinetic patterns in embryos derived from obese and normoweight infertile women: a time-lapse study. *Hum Reprod* 2013;**28**:794–800.

Bodri D, Sugimoto T, Serna JY, Kondo M, Kato R, Kawachiya S, Matsumoto T. Influence of different oocyte insemination techniques on early and late morphokinetic parameters: retrospective analysis of 500 time-lapse monitored blastocysts. *Fertil Steril* 2015;**104**:1175–1181.

Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Hickman CF. Modelling a risk classification of aneuploidy in human embryos using non-invasive morphokinetics. *Reprod Biomed Online* 2013a;**26**:477–485.

Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reprod Biomed Online* 2013b;**27**:140–146.

Ciray HN, Aksoy T, Goktas C, Ozturk B, Bahceci M. Time-lapse evaluation of human embryo development in single versus sequential culture media—a sibling oocyte study. *J Assist Reprod Genet* 2012;**29**:891–900.

Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, Baker VL, Adamson GD, Abusief ME, Gvakharia M et al. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril* 2013;**100**:412–419 e415.

Cruz M, Garrido N, Herrero J, Perez-Cano I, Munoz M, Meseguer M. Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reprod Biomed Online* 2012;**25**:371–381.

Cruz M, Garrido N, Gadea B, Munoz M, Perez-Cano I, Meseguer M. Oocyte insemination techniques are related to alterations of embryo developmental timing in an oocyte donation model. *Reprod Biomed Online* 2013;**27**:367–375.

Dal Canto M, Coticchio G, Mignini Renzini M, De Ponti E, Novara PV, Brambillasca F, Comi R, Fadini R. Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation. *Reprod Biomed Online* 2012;**25**:474–480.

Freour T, Dessolle L, Lammers J, Lattes S, Barriere P. Comparison of embryo morphokinetics after in vitro fertilization-intracytoplasmic sperm injection in smoking and nonsmoking women. *Fertil Steril* 2013;**99**:1944–1950.

Freour T, Le Fleuter N, Lammers J, Splingart C, Reignier A, Barriere P. External validation of a time-lapse prediction model. *Fertil Steril* 2015;**103**:917–922.

Kaser DJ, Racowsky C. Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review. *Hum Reprod Update* 2014;**20**:617–631.

Kirkegaard K, Agerholm IE, Ingerslev HJ. Time-lapse monitoring as a tool for clinical embryo assessment. *Hum Reprod* 2012a;**27**:1277–1285.

Kirkegaard K, Hindkjaer JJ, Grondahl ML, Kesmodel US, Ingerslev HJ. A randomized clinical trial comparing embryo culture in a conventional incubator with a time-lapse incubator. *J Assist Reprod Genet* 2012b;**29**:565–572.

Kirkegaard K, Hindkjaer JJ, Ingerslev HJ. Effect of oxygen concentration on human embryo development evaluated by time-lapse monitoring. *Fertil Steril* 2013a;**99**:738–744 e734.

Kirkegaard K, Kesmodel US, Hindkjaer JJ, Ingerslev HJ. Time-lapse parameters as predictors of blastocyst development and pregnancy outcome in embryos from good prognosis patients: a prospective cohort study. *Hum Reprod* 2013b;**28**:2643–2651.

Kirkegaard K, Campbell A, Agerholm I, Bentin-Ley U, Gabrielsen A, Kirk J, Sayed S, Ingerslev HJ. Limitations of a time-lapse blastocyst prediction model: a large multicentre outcome analysis. *Reprod Biomed Online* 2014a;**29**:156–158.

Kirkegaard K, Svane AS, Nielsen JS, Hindkjaer JJ, Nielsen NC, Ingerslev HJ. Nuclear magnetic resonance metabolomic profiling of Day 3 and 5 embryo culture medium does not predict pregnancy outcome in good prognosis patients: a prospective cohort study on single transferred embryos. *Hum Reprod* 2014b;**29**:2413–2420.

Kirkegaard K, Ahlstrom A, Ingerslev HJ, Hardarson T. Choosing the best embryo by time lapse versus standard morphology. *Fertil Steril* 2015;**103**:323–332.

Kirkwood B, Sterne J. *Essential Medical Statistics,* 2nd edn. United Kingdom: John Wiley and Sons Ltd, 2003.

Kokkali G, Vrettou C, Traeger-Synodinos J, Jones GM, Cram DS, Stavrou D, Trounson AO, Kanavakis E, Pantos K. Birth of a healthy infant following trophectoderm biopsy from blastocysts for PGD of beta-thalassaemia major. *Hum Reprod* 2005;**20**:1855–1859.

Meseguer M, Herrero J, Tejera A, Hilligsoe KM, Ramsing NB, Remohi J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod* 2011;**26**:2658–2671.

Meseguer M, Rubio I, Cruz M, Basile N, Marcos J, Requena A. Embryo incubation and selection in a time-lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. *Fertil Steril* 2012;**98**:1481–1489 e1410.

Munoz M, Cruz M, Humaidan P, Garrido N, Perez-Cano I, Meseguer M. Dose of recombinant FSH and oestradiol concentration on day of HCG affect embryo development kinetics. *Reprod Biomed Online* 2012; **25**:382–389.

Munoz M, Cruz M, Humaidan P, Garrido N, Perez-Cano I, Meseguer M. The type of GnRH analogue used during controlled ovarian stimulation influences early embryo developmental kinetics: a time-lapse study. *Eur J Obstet Gynecol Reprod Biol* 2013;**168**:167–172.

Ottolini C, Rienzi L, Capalbo A. A cautionary note against embryo aneuploidy risk assessment using time-lapse imaging. *Reprod Biomed Online* 2014; **28**:273–275.

Rienzi L, Capalbo A, Stoppa M, Romano S, Maggiulli R, Albricci L, Scarica C, Farcomeni A, Vajta G, Ubaldi FM. No evidence of association between blastocyst aneuploidy and morphokinetic assessment in a selected population of poor-prognosis patients: a longitudinal cohort study. *Reprod Biomed Online* 2015;**30**:57–66.

Rubio I, Galan A, Larreategui Z, Ayerdi F, Bellver J, Herrero J, Meseguer M. Clinical validation of embryo culture and selection by morphokinetic analysis: a randomized, controlled trial of the EmbryoScope. *Fertil Steril* 2014;**102**:1287–1294.

Storr A, Venetis CA, Cooke S, Susetio D, Kilani S, Ledger W. Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. *J Assist Reprod Genet* 2015;**32**:1151–1160.

Sundvall L, Ingerslev HJ, Breth Knudsen U, Kirkegaard K. Inter- and intra-observer variability of time-lapse annotations. *Hum Reprod* 2013; **28**:3215–3221.

Sundvall L, Kirkegaard K, Ingerslev HJ, Knudsen UB. Unaltered timing of embryo development in women with polycystic ovarian syndrome (PCOS): a time-lapse study. *J Assist Reprod Genet* 2015;**32**:1031–1042.

van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;**16**:577–589.

VerMilyea MD, Tan L, Anthony JT, Conaghan J, Ivani K, Gvakharia M, Boostanfar R, Baker VL, Suraj V, Chen AA et al. Computer-automated time-lapse analysis results correlate with embryo implantation and clinical pregnancy: A blinded, multi-centre study. *Reprod Biomed Online* 2014;**29**:729–736.