Between-laboratory reproducibility of time-lapse embryo selection using qualitative and quantitative parameters: a systematic review and meta-analysis

Yanhe Liu^{1,2,3} · Fang Qi⁴ · Phillip Matson² · Dean E. Morbeck⁵ · Ben W. Mol⁶ · Sai Zhao⁴ · Masoud Afnan⁷

Received: 17 December 2019 / Accepted: 17 April 2020 / Published online: 2 May 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose To investigate the between-laboratory reproducibility of embryo selection/deselection effectiveness using qualitative and quantitative time-lapse parameters.

Methods A systematic search was performed on MEDLINE, EMBASE, and the Cochrane Library (up to February 2020) without restriction on date, language, document type, and publication status. Measuring outcomes included implantation, blastulation, good-quality blastocyst formation, and euploid blastocyst.

Results We detected 6 retrospective cohort studies externally validating the first clinical time-lapse model (Meseguer) emphasizing quantitative parameters, of which 3 (including one involving 2 independent centers) were included for the pooled analysis. Receiver operating characteristics analysis showed reduced predictive power of the model when either including or not including sister clinic validation. Fifteen cohort studies evaluating qualitative parameters were included for meta-analysis, and the mean Newcastle-Ottawa Scale was 5.3. Overall, meta-analysis showed significantly adverse association between the presence of ≥ 1 cleavage abnormalities and embryo implantation rates (11 studies, $n = 7266$; $RR = 0.39 [0.28, 0.55]_{95\% CI}$; $I^2 = 57\%$). Further analysis showed adverse impacts of direct cleavage (7 studies, $n = 7065$; $RR = 0.28 [0.15, 0.54]_{95\% CI}$; $I^2 = 46\%$), reverse cleavage (2 studies, $n = 3622$; $RR = 0.16 [0.03, 0.75]_{95\% CI}$; $I^2 = 0\%$), chaotic cleavage (2 studies, $n = 3643$; $RR = 0.11 [0.02, 0.69]_{95\% CI}$; $I^2 = 24\%$), and multinucleation (5 studies, $n = 2576$; $RR = 0.59 [0.50, 0.69]_{95\% CI}$; $I^2 = 0\%$), but not the < 6 intercellular contact points at the 4-cell stage (1 study, $n = 185$; $RR = 0.17 [0.02, 1.15]_{95\% CI}$).

Conclusions Qualitative time-lapse parameters are reliably associated with embryo developmental potential among laboratories, whereas the reproducibility of time-lapse embryo selection model that emphasizes quantitative parameters may be compromised when externally applied.

Keywords Time-lapse · Embryo selection · In vitro fertilization · Meta-analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10815-020-01789-4>) contains supplementary material, which is available to authorized users.

✉ Yanhe Liu
gift0409@yahoo.com.au

¹ Reproductive Medicine Center, Tianjin United Family Hospital, Tianjin, China

² School of Medical and Health Sciences, Edith Cowan University, Joondalup, Australia

³ School of Human Sciences, The University of Western Australia, Crawley, Australia

⁴ Systematic Review Solutions Ltd, Shanghai, China

⁵ Fertility Associates, Auckland, New Zealand

⁶ Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia

⁷ Department of Obstetrics and Gynaecology, Qingdao United Family Hospital, Qingdao, China

Introduction

Following the introduction of time-lapse (TL) videography into assisted reproductive technology (ART) laboratories in the early 2010s [1], growing interest has been evident among embryologists globally, with numerous attempts to improve embryo selection by using such technology [2–5]. Indeed, TL incubation not only provides human embryos with uninterrupted culture conditions, but also offers embryologists the novel opportunity in improving embryo selection based on the application of morphokinetic algorithms. Although the evidence for transferability of existing TL embryo selection models is inconclusive in recent years, questioning of the concept that one single TL model has universal application across different clinics is being raised [6–8]. This lack of broad applicability is likely because morphokinetics of *in vitro* cultured embryos are significantly affected by varying laboratory culture conditions along with different patient characteristics [9, 10]. To counter inter-clinic heterogeneity, TL model with a larger dataset including multicentric embryos was developed [5]; however, this model was not successfully validated with an independent dataset [7]. Transferability issues may lead to overestimation by users regarding the potential benefits of any specific TL model, in particular when commercial promotions are involved. Current evidence does not support total reliance on either externally published or commercially purchased TL models before further *in house* validation, although the option of adopting published methodology to develop individualized model appears to be practical [3].

Apart from quantitative TL measures (i.e., morphokinetics such as time to reach the 5-cell stage) which are mostly continuous parameters in the statistical perspective, new TL measures that go largely undetectable via conventional static microscopic observations are qualitative parameters. Examples of these parameters include (i) direct cleavage (DC) where one blastomere directly divides into three or more daughter blastomeres [11], (ii) reverse cleavage (RC) where blastomeres either fuse after cleavage or lead to failed cytokinesis following karyokinesis [12], (iii) chaotic cleavage (CC) where cell cleavage display erratic and disordered behavior [13], and (iv) embryos with less than 6 intercellular contact points (ICCP) at the 4-cell stage [14]. The assessment of certain conventional measures such as multinucleation (MN) at the 2- or 4-cell stage could also be improved due to more complete dynamic information obtained via TL monitoring [15]. The use of these parameters is independent of absolute cleavage timings of embryos; therefore, they are not likely associated with culture conditions or patient characteristics. This systematic review aims to summarize all external validation studies of the first clinically reported TL model emphasizing quantitative parameters, followed by evaluation via pooled receiver operating characteristics (ROC) analysis. Furthermore, studies reporting the impact of qualitative TL measures on the developmental and implantation potential of embryos are systematically reviewed followed by meta-analysis.

Methods

Protocol for this review was registered on the international prospective register of systematic reviews (PROSPERO, registration number CRD42018098826).

Search strategies and selection criteria

We searched MEDLINE, EMBASE, and the Cochrane Library (up to February 2020) without restriction on date, language, document type, and publication status. Detailed search strategies were presented in the Supplemental File 1. The full list of records hit and associated systematic review articles were crosschecked, with the aim to include as many eligible studies as possible. Two reviewing authors (FQ and SZ) independently screened all references, and full-text publications were included when the following criteria were met: (i) comparative studies including randomized and non-randomized controlled trials (RCTs and NCTs), observational cohort studies and case-control studies; (ii) all women, regardless of age or number of previous attempts, undergoing *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI) treatment with the use of TL incubation for resulting embryos; (iii) studies investigating the developmental or implantation potential of embryos with any type of abnormal biological events observed via TL videography up to the 8-cell stage. Any disagreement between the 2 reviewers was resolved via judgment by a third author (YL) as required.

Outcomes of interest

The main outcome was successful implantation, defined by detection of intrauterine fetal heart beat under ultrasound 5–7 weeks post egg collection. The secondary outcomes included blastocyst formation, defined by whether or not embryo having reached at least stage 2 (as per the Gardner's grade) 5 to 6 days post egg collection [16]; good-quality blastocyst rate, defined by whether or not embryo having reached at least 3BB (as per the Gardner's grade) 5 to 6 days post egg collection [16]; and ploidy status of embryos (euploidy or not) if preimplantation genetic screening was performed on observed embryos.

Data extraction and assessment for risk of bias

Two reviewing authors (FQ and SZ) independently extracted quantitative and qualitative information by using a standard data extraction form. In brief, the Cochrane tool for Risk of Bias assessment was chosen to evaluate every domain of risk of bias for RCTs [17]. For NCTs and observational studies (cohort studies and case-control studies), the Newcastle-Ottawa Scale (NOS) tool was used to evaluate items considered relevant to this review topic [18]. Any disagreement between the 2 reviewing authors was resolved via judgment by a third author (YL) as required.

Statistical analysis for quantitative parameter-based embryo selection model

One of the first published and most cited TL algorithms for embryo selection, the Meseguer [1] model, was selected for evaluating its transferability when subsequently validated by other laboratories. ROC analysis was performed and area under the curve (AUC) along with 95% confidence intervals (CIs) was used to evaluate this model's predictive power on successful implantation of embryos, via pooled analysis (R software, [19]).

Statistical analysis for qualitative parameters

All predefined outcomes were dichotomous, and the effects were presented by using summarized risk ratios (RRs) with 95% CI, with random-effect model being used for all meta-analyses in the R software [19]. The clinical and methodological heterogeneity of included studies were thoroughly assessed, with an $I^2 > 50\%$ coupled with $P < 0.1$ following chi-squared test considered evidence of substantial levels of statistical heterogeneity.

Additional analysis

Subgroup analysis was performed on the basis of the following characteristics: (i) MN at the 2- or 4-cell stage and (ii) abnormal cleavage patterns, including DC (or tripolar cleavage/division), type I RC (or cell fusion), type II RC (or failed/no cleavage/division), less than 6 ICCPs at the 4-cell stage (or planar/non-tetrahedral/one-plate embryos), and CC (or chaotic division).

Analyses or additional sensitivity analyses were not carried out for further subgrouping due to insufficient data or wide CIs for most treatment estimates. Reporting bias was assessed through visual interpretation of funnel plots only for outcomes including 10 or more studies (Supplemental file 1).

Results

Studies included for quantitative assessment

After removal of duplicates, a total of 544 records were hit with 6 eligible studies identified. Study selection process is shown in PRISMA flow diagram (Fig. 1a). All 6 studies were retrospective cohort studies, performed in a range of countries from different continents, involving either a single center or multiple centers with varying sample sizes. Three out of the 6 studies were excluded from pooled analysis due to either (i) being the original report of the target model [1], or (ii) not reporting the total number in each embryo category [6], or (iii) the use of blastocyst formation instead of implantation as the end point [5]. Detailed study characteristics are summarized in Supplemental Table 1.

Pooled analysis on implantation

Three studies [8, 14, 20] included numerators and denominators for calculation of implantation rates in different grading categories according to the Meseguer model [1] (Fig. 2). In comparison to the originally reported predictive power of the model ($n = 247$; $AUC = 0.720[0.645–0.795]_{95\% \text{ CI}}$), ROC

Fig. 1 PRISMA flow diagram. **a** Quantitative assessment. **b** Qualitative assessment.

Fig. 2 The predictive power (pooled receiver operating characteristics analysis) for implantation

analysis after pooling data from 3 external validation studies showed a reduction in AUC ($n = 1667$; AUC = 0.567[0.538–0.597]_{95% CI}), although still statistically significant. However, after removal of the Rubio study [20] from the pooled analysis (if regarded as an internal validation since both laboratories are associated within the same group), the predictive power of the Meseguer model [1] decreased and lost statistical significance ($n = 1155$; AUC = 0.510[0.473–0.547]_{95% CI}) (Fig. 2).

Studies included for qualitative assessment

After removal of duplicates, a total of 599 records were hit with 20 eligible studies identified. Study selection process is shown in PRISMA flow diagram (Fig. 1b). All 20 studies were cohort studies focusing on embryos with or without abnormal biological events, involving either a single institute or multiple institutes. Studies were conducted in various regions including Asia (China, Israel, Japan, Korea, Turkey), Europe (Denmark, France, Italy, Slovenia, Spain, UK), North America (Canada, USA), and Oceania (Australia). Using NOS, all included studies were of moderate quality with a mean score of 5.3. Most studies (90%) were retrospective and lacked descriptive statements on their confounding control and follow-up protocols, suggesting a potential bias in comparability between groups. Among the 20 eligible studies, 15 [2, 11–15, 21–29] were eventually included for meta-

analysis with the other 5 [21, 30–33] described separately due to unavailability of raw data for merging. The study characteristics and quality assessment of all included studies were summarized in Supplemental Table 2.

Meta-analysis for qualitative assessment

Implantation

Eleven included studies [11–15, 22, 23, 25, 27–29] used implantation as the end point to study overall impacts of any abnormal biological event of embryos. Results indicated that embryos with the presence of any abnormal biological event had lower implantation rates than embryos that lacked an abnormal event (11 studies, $n = 7266$; RR = 0.39[0.28, 0.55]_{95% CI}; $I^2 = 57%$; Fig. 3). There was moderate to substantial heterogeneity for this outcome. Further, we conducted subgroup analysis based on different types of abnormal biological events. The results revealed adverse impacts of abnormal biological events on the implantation outcome, including DC (7 studies, $n = 7065$; RR = 0.28[0.15, 0.54]_{95% CI}; $I^2 = 46%$) [2, 11, 13, 23, 25, 26, 28], RC (2 studies, $n = 3622$, RR = 0.16[0.03, 0.75]_{95% CI}; $I^2 = 0%$) [12, 23], CC (2 studies, $n = 3643$, RR = 0.11[0.02, 0.69]_{95% CI}; $I^2 = 24%$) [13, 23], and MN (5 studies, $n = 2576$, RR = 0.59[0.50, 0.69]_{95% CI}; $I^2 = 0%$) [2, 15, 22, 26, 28]. The feature < 6 ICCPs (1 study, $n =$

Fig. 3 Forest plots of implantation outcomes of embryos with or without any abnormal biological event

185, $RR = 0.17[0.02, 1.15]_{95\% \text{ CI}}$ [14] was not associated with implantation outcome (Fig. 4). Desch et al. [21] also confirmed a significantly negative effect of MN at the 2-cell ($OR = 0.57[0.34, 0.94]_{95\% \text{ CI}}$) and the 4-cell stages ($OR = 0.51[0.28, 0.94]_{95\% \text{ CI}}$) separately.

Blastulation

Four studies [13, 24, 29, 34] were included for meta-analysis to investigate overall impact of any abnormal biological event of embryos on the subsequent blastulation. Results showed lower blastulation rate in embryos with the presence of any abnormal biological event in comparison to their normal counterparts (4 studies, $n = 2457$; $RR = 0.59[0.49, 0.71]_{95\% \text{ CI}}$; $I^2 = 73\%$). There was, however, high heterogeneity among the 4 studies. Further subgroup analysis based on different patterns of abnormal biological events showed clear adverse impact of DC on subsequent blastulation [13, 24, 26] (3 studies, $n = 1619$; $RR = 0.48[0.34, 0.66]_{95\% \text{ CI}}$; $I^2 = 55\%$) but not the other patterns (RC and MN).

Good-quality blastocyst formation

Six studies [13, 26, 27, 29, 31, 32] reported data using good-quality blastocyst formation as an end point but with very high statistical heterogeneity ($> 90\%$), hence the unsuitability to perform meta-analysis. Yang et al. [29] did not describe the specific type of cleavage abnormalities and found that embryos with the presence of abnormal biological event had lower chance to form good-quality blastocysts ($n = 345$; $RR = 0.38[0.30, 0.48]_{95\% \text{ CI}}$). Hojnik et al. [27] reported similar results ($n = 53$; $RR = 0.03[0.00, 0.46]_{95\% \text{ CI}}$). Four studies [13, 26, 31, 32] reported the impact of DC on subsequent chance of embryos to form good-quality blastocysts. Among these, the majority [26, 31,

32] identified adverse impact of DC on good-quality blastocyst formation but one did not [13]. Two studies [26, 31] also investigated the impact of MN but did not detect any impact on good-quality blastocyst formation rate.

Ploidy status

Two studies [24, 34] reported data using ploidy status as an end point. No difference was identified in the proportion of euploid blastocysts between embryos with and without any abnormal biological event (2 studies, $n = 1279$; $RR = 0.74[0.54, 1.01]_{95\% \text{ CI}}$). Similar results were found when assessing the impact of DC and RC on subsequent ploidy outcomes.

Discussion

As TL embryo selection gains in popularity, different meta-analyses have failed to gain consensus on the advantage of TL over conventional culture/selection [35, 36]. In 2011, Meseguer et al. [1] published the first clinical algorithm for improved embryo selection via the use of TL incubation, which opened a new era in embryo selection as one of the critical parts of IVF treatment. With additional models being subsequently reported by different researchers, external validation of the Meseguer model was inconclusive, with both negative [6, 8, 37] and positive reports [20]. More recently, published work also revealed poor transferability when several models were applied on the same external dataset, including a commercially available model which had been claimed in the original report to be universally applicable [7]. Commercial tools such as the Kidscore use predominantly quantitative measures and are presumably limited by having large reference ranges to fit multiple laboratories. The current

Fig. 4 Forest plots of implantation outcomes of embryos with or without specific abnormal biological event. **a** Direct cleavage. **b** Multinucleation. **c** Reverse cleavage. **d** Chaotic cleavage. **e** < 6 intercellular contact points at the 4-cell stage

study confirms this lack of universality as it clearly demonstrates the reduction in predictive power (AUC from 0.720 to 0.510) after pooling data from different studies. Indeed, the transferability issue was mainly due to different morphokinetic profiles of embryos (i.e., quantitative parameters) between laboratories [6–8, 37], where the culture conditions employed (such as type of culture media or oxygen concentration in the incubation gas phase) could significantly alter embryo's developmental speed ([9, 38]). Therefore, clinics within the same institution are more likely to achieve consistent results [20].

The findings presented here suggest a change in the current strategy employed for ranking embryos when using timing-based quantitative measures. For instance, rather than defining an “optimal” range of timing (which varies between laboratories), a strategy of picking the faster ones (which does not vary between laboratories) may improve generalizability. However, this would require the removal from analysis of embryos displaying DC as proposed earlier [3], as these embryos divide at a higher rate but with much reduced implantation potential [11]. This strategy is therefore considered helpful in improving transferability of TL embryo selection models [3]. The identification of DC embryos is generally done through either (i) manual annotation of tripolar division with 3 resulting nucleated daughter cells or (ii) a simple calculation for a short 2-cell stage (<5 h) [39]. However, the latter approach is limited to identifying DC at the 1-cell stage and is not applicable when detecting DC in blastomeres at the 2-cell or 4-cell stage. Furthermore, the presence of a nucleus in each daughter cell following any tripolar division may not be always confirmed by embryologists in daily practice (as per in house training experience on TL annotation), due to high workload or poor viewing angle of images. Computer-assisted identification of DC is considered even less reliable as nuclear detection is still currently unachievable by most TL software. The detection of later stage DC events is also difficult to achieve via calculations like “2-cell stage < 5h” for the first cell division post fertilization, especially when DC events occur in multiple blastomeres of the same embryo. Similarly, the identification of RC also requires confirmation of karyokinesis along with cytokinesis, which applies to both type I (cell fusion) and Type II (no cleavage after karyokinesis) as originally reported [12]. Therefore, large-scale promotions of TL algorithms that are driven by commercialization seem less in favor of such difficult and time-consuming annotation of qualitative parameters, and most commercially available TL algorithms appeared to be obliged to use quantitative strategies as a result. Indeed, the identification of qualitative TL features is considered difficult for embryologists as well, which usually requires extensive training supported by ongoing internal quality control and ideally external quality assurance programs. This however has rarely been discussed either in literature or the commercial TL promotion activities.

It has been recently proposed that qualitative parameters may improve reproducibility of embryo deselection between different laboratories [40]. Binary outcomes of qualitative parameters are independent of absolute cleavage timings of embryos. In contrast to qualitative parameters, a study in two unrelated clinics found that even implanting embryos grew at very different rates, with the difference in time to reach the 8-cell stage (relative to pronuclear fading) being greater than 3 h [37]. This may imply the ability of in vitro cultured human embryos to adapt in response to different culture conditions without losing their implantation potential. The present meta-analysis focused on DC, RC, CC, < 6 ICCP, and MN, with the primary outcome of implantation and secondary outcomes of blastulation and ploidy status. Based on these findings, future TL models built upon a combination of different qualitative parameters would likely show advantages over quantitative parameters in terms of reproducibility. Additional qualitative parameters that were not included by the present meta-analysis could potentially contribute to a continuously growing matrix for further advanced embryo deselection.

One limitation of the current study is the use of implantation as primary outcome. Indeed, live birth is the widely accepted optimal end point for IVF studies. But in consideration of the small number of TL studies using live birth as an end point when performing the current meta-analysis, we used implantation as the surrogate end point so as to improve statistical power. Furthermore, RCTs are the preferred type of studies for meta-analysis assessment considering their high-quality evidence over retrospective studies. Studies included in the current meta-analysis are mostly retrospective, with quality of evidence graded from very low to moderate, as well as a lack in descriptive statements on their confounding control and follow-up protocols. Therefore, these studies are subject to potentially increased bias in comparability between groups. Besides, the various embryo culture conditions applied in studies included for this meta-analysis led to reduced homogeneity for combined analysis. Nonetheless, the results showed high consistency in outcomes among different laboratories with a consistent finding of an adverse impact of qualitative parameters. The nature of these abnormal developmental characteristics in fact makes RCTs unlikely, given potential ethical issues (i.e., randomizing participants into the test group where such obvious deselection biomarkers are ignored from assessment). In addition, the consistently adverse impacts of studied qualitative parameters may, to a certain extent, be affected by publication bias, where positive results are more easily published than the negative ones. Future multi-centric RCTs comparing quantitative parameter-based embryo selection and qualitative parameter-based deselection are necessary as the TL deselection approach may have been overlooked and underestimated. This could also lead to more effective and transferable way for TL embryo selection.

It is worth pointing out that the observed compromised reproducibility of the Meseguer model when externally applied may not necessarily represent more complex models, such as those developed through regression analysis using ideology-based approaches [41]. Additional meta-analysis on such more recent models is desirable when further external validation data become available in literature. Meanwhile, the rapid progress of artificial intelligence coupled with TL imaging technology may shed new light on fresh strategies for machine learning-oriented embryo selection. Encouragingly, a recent machine learning program was reported to be able to predict fetal heart at a revolutionarily high level of AUC (> 0.9), but as pointed out by the authors, this study could not specify the exact logic employed by the computer when developing the algorithm rather than “feeding the machine with embryo videos” [42]. Such a high level of AUC requires careful external validation by independent institutes prior to its clinical introduction. In the meantime, a basic biological understanding of the “black box” process by embryologists may be considered ethically necessary [43], since concerns about machine learning algorithms are basically hidden and therefore are not transparent and not subject to human oversight or peer review.

Conclusions

Qualitative TL parameters are reproducible embryo deselection criteria for implantation potential, while quantitative TL parameter-based models are less reliable when introduced into an external laboratory.

References

- 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;26:2658–71.
- Basile N, Vime P, Florensa M, Aparicio Ruiz B, Garcia Velasco JA, Remohi J, et al. The use of morphokinetics as a predictor of implantation: a multicentric study to define and validate an algorithm for embryo selection. *Hum Reprod*. 2015;30:276–83.
- Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Time-lapse deselection model for human day 3 in vitro fertilization embryos: the combination of qualitative and quantitative measures of embryo growth. *Fertil Steril*. 2016;105:656–62 e651.
- Milewski R, Kuc P, Kuczynska A, Stankiewicz B, Lukaszuk K, Kuczynski W. A predictive model for blastocyst formation based on morphokinetic parameters in time-lapse monitoring of embryo development. *J Assist Reprod Genet*. 2015;32:571–9.
- Petersen BM, Boel M, Montag M, Gardner DK. Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on day 3. *Hum Reprod*. 2016;31:2231–44.
- Freour T, Le Fleuter N, Lammers J, Splingart C, Reignier A, Barriere P. External validation of a time-lapse prediction model. *Fertil Steril*. 2015;103:917–22.
- Liu Y, Feenan K, Chapple V, Matson P. Assessing efficacy of day 3 embryo time-lapse algorithms retrospectively: impacts of dataset type and confounding factors. *Hum Fertil (Camb)*. 2019;22:182–90.
- Yalcinkaya E, Ergin EG, Caliskan E, Oztel Z, Ozay A, Ozornek H. Reproducibility of a time-lapse embryo selection model based on morphokinetic data in a sequential culture media setting. *J Turk Ger Gynecol Assoc*. 2014;15:156–60.
- Kirkegaard K, Hindkjaer JJ, Ingerslev HJ. Effect of oxygen concentration on human embryo development evaluated by time-lapse monitoring. *Fertil Steril*. 2013;99:738–44 e734.
- Wissing ML, Bjerre MR, Olesen AIG, Hoest T, Mikkelsen AL. Impact of PCOS on early embryo cleavage kinetics. *Reprod BioMed Online*. 2014;28:508–14.
- Rubio I, Kuhlmann R, Agerholm I, Kirk J, Herrero J, Escriba MJ, et al. Limited implantation success of direct-cleaved human zygotes: a time-lapse study. *Fertil Steril*. 2012;98:1458–63.
- Liu Y, Chapple V, Roberts P, Matson P. Prevalence, consequence, and significance of reverse cleavage by human embryos viewed with the use of the Embryoscope time-lapse video system. *Fertil Steril*. 2014;102:1295–1300.e1292.
- Athayde Wirka K, Chen AA, Conaghan J, Ivani K, Gvakharia M, Behr B, et al. Atypical embryo phenotypes identified by time-lapse microscopy: high prevalence and association with embryo development. *Fertil Steril*. 2014; 1637-1648(101):e1631–5.
- Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Clinical significance of intercellular contact at the four-cell stage of human embryos, and the use of abnormal cleavage patterns to identify embryos with low implantation potential: a time-lapse study. *Fertil Steril*. 2015;103:1485–1491.e1481.
- Ergin EG, Caliskan E, Yalcinkaya E, Oztel Z, Cokelez K, Ozay A, et al. Frequency of embryo multinucleation detected by time-lapse system and its impact on pregnancy outcome. *Fertil Steril*. 2014;102:1029–1033.e1021.
- Gardner DK, Lane M, Stevens J, Schlenker T, Schoolcraft WB. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil Steril*. 2000;73:1155–8.
- Higgins JPT GSe. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011.
- Wells GASB, O’Connell D, Peterson J, Welch V, et al. The Newcastle -Ottawa scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. 2011. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 19 Oct 2019
- Team RC.R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017.
- Rubio I, Galan A, Larreategui Z, Ayerdi F, Bellver J, Herrero J, et al. Clinical validation of embryo culture and selection by morphokinetic analysis: a randomized, controlled trial of the EmbryoScope. *Fertil Steril*. 2014;102:1287–1294.e1285.
- Desch L, Bruno C, Luu M, Barberet J, Choux C, Lamotte M, et al. Embryo multinucleation at the two-cell stage is an independent predictor of intracytoplasmic sperm injection outcomes. *Fertil Steril*. 2017;107:97–103.e104.
- Balakier H, Sojecki A, Motamedi G, Librach C. Impact of multinucleated blastomeres on embryo developmental competence, morphokinetics, and aneuploidy. *Fertil Steril*. 2016;106:608–614.e602.
- Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. Preliminary investigation of the prevalence and implantation potential of abnormal embryonic phenotypes assessed using time-lapse imaging. *Reprod BioMed Online*. 2017;34:455–62.

24. Desai N, Goldberg JM, Austin C, Falcone T. Are cleavage anomalies, multinucleation, or specific cell cycle kinetics observed with time-lapse imaging predictive of embryo developmental capacity or ploidy? *Fertil Steril*. 2018;109:665–74.
25. Fan YL, Han SB, Wu LH, Wang YP, Huang GN. Abnormally cleaving embryos are able to produce live births: a time-lapse study. *J Assist Reprod Genet*. 2016;33:379–85.
26. Hashimoto S, Nakano T, Yamagata K, Inoue M, Morimoto Y, Nakaoka Y. Multinucleation per se is not always sufficient as a marker of abnormality to decide against transferring human embryos. *Fertil Steril*. 2016;106:133–139.e136.
27. Hojnik N, Vlasisavljevic V, Kovacic B. Morphokinetic characteristics and developmental potential of in vitro cultured embryos from natural cycles in patients with poor ovarian response. *Biomed Res Int*. 2016;2016:4286528.
28. Hur YS, Ryu EK, Hyun CS, Yang SH, Yoon SH, Lim KS, et al. Retrospective study of single vitrified-warmed blastocyst transfer cycles according to the presence of morphokinetic variables. *Clin Exp Reprod Med*. 2018;45:52–5.
29. Yang ST, Shi JX, Gong F, Zhang SP, Lu CF, Tan K, et al. Cleavage pattern predicts developmental potential of day 3 human embryos produced by IVF. *Reprod BioMed Online*. 2015;30:625–34.
30. Almagor M, Or Y, Fieldust S, Shoham Z. Irregular cleavage of early preimplantation human embryos: characteristics of patients and pregnancy outcomes. *J Assist Reprod Genet*. 2015;32:1811–5.
31. 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*. 2013;28:2643–51.
32. Kong X, Yang S, Gong F, Lu C, Zhang S, Lu G, et al. The relationship between cell number, division behavior and developmental potential of cleavage stage human embryos: a time-lapse study. *PLoS ONE [Electronic Resource]*. 2016;11:e0153697.
33. Yang SH, Wu CH, Chen YC, Yang CK, Wu TH, Chen PC, et al. Effect of morphokinetics and morphological dynamics of cleavage stage on embryo developmental potential: a time-lapse study. *Taiwan J Obstet Gynecol*. 2018;57:76–82.
34. Lagalla C, Tarozzi N, Sciajno R, Wells D, Di Santo M, Nadalini M, et al. Embryos with morphokinetic abnormalities may develop into euploid blastocysts. *Reprod BioMed Online*. 2017;34:137–46.
35. Armstrong S, Bhide P, Jordan V, Pacey A, Farquhar C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev*. 2018;5:CD011320.
36. Pribenszky C, Nilselid AM, Montag M. Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis. *Reprod BioMed Online*. 2017;35:511–20.
37. Liu Y, Copeland C, Stevens A, Feenan K, Chapple V, Myssonski K, et al. Assessment of human embryos by time-lapse videography: a comparison of quantitative and qualitative measures between two independent laboratories. *Reprod Biol*. 2015;15:210–6.
38. 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.
39. Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, et al. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod*. 2014;29:2650–60.
40. Liu Y, Sakkas D, Afnan M, Matson P. Time-lapse videography for embryo selection/de-selection: a bright future or fading star? *Hum Fertil (Camb)*. 2019; 9:1–7.
41. Fishel S, Campbell A, Montgomery S, Smith R, Nice L, Duffy S, et al. Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth. *Reprod BioMed Online*. 2018;37:304–13.
42. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod*. 2019;34:1011–8.
43. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018;15:e1002689.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.