ORIGINAL ARTICLE: ASSISTED REPRODUCTION

Time-lapse algorithms and morphological selection of day-5 embryos for transfer: a preclinical validation study

Ashleigh Storr, Pg.D., Christos Venetis, Ph.D., Simon Cooke, Ph.D., Suha Kilani, Ph.D., and William Ledger, Ph.D.

IVF Australia, Sydney; and School of Women's and Children's Health, Faculty of Medicine, University of New South Wales, Randwick, New South Wales, Australia

Objective: To determine the agreement between published time-lapse algorithms in selecting the best day-5 embryo for transfer, as well as the agreement between these algorithms and embryologists.
Design: Prospective study.
Setting: Private in vitro fertilization center.
Patient(s): Four hundred and twenty-eight embryos from 100 cycles cultured in the EmbryoScope.
Intervention(s): None.
Main Outcome Measure(s): Interalgorithm agreement as assessed by the Fleiss kappa coefficient.
Result(s): Of seven published algorithms analyzed in this study, only one of the 18 possible pairs showed very good agreement ($\kappa = 0.867$); one pair showed good agreement ($\kappa = 0.725$), four pairs showed fair agreement ($\kappa = 0.226$–0.334), and the remaining 12 pairs showed poor agreement ($\kappa = 0.008$–0.149). Even in the best-case scenario, the majority of algorithms showed poor to moderate kappa scores ($\kappa = 0.337$–0.722) for the assessment of agreement between the embryo(s) selected as "best" by the algorithms and the embryo that was chosen by the majority (>5) of embryologists, as well as with the embryo that was actually selected in the laboratory on the day of transfer ($\kappa = 0.315$–0.802).
Conclusion(s): The results of this study raise concerns as to whether the tested algorithms are applicable in different clinical settings, emphasizing the need for proper external validation before clinical use. (Fertil Steril® 2017;■:■–■. ©2017 by American Society for Reproductive Medicine.)
Key Words: Agreement, algorithm, embryologist, morphokinetics, time-lapse

Discuss: You can discuss this article with its authors and other ASRM members at https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/22076-24320

The selection of the best embryo for transfer has remained dependent on the morphological assessment of embryos since the introduction of assisted reproductive technology (1–3). This process, however, involves the understanding and application of detailed grading methods by individual embryologists, which has previously been shown to be highly subjective (4–7). Because of this, some variation in embryo scoring and selection is anticipated among different assessors, potentially affecting the accuracy of morphology-based embryo selection in choosing the embryo with presumably the highest implantation potential (3, 4, 7).

The introduction of time-lapse incubators has allowed for the uninterrupted culture of embryos and the provision of significantly more information to the embryologist regarding embryo growth and development. Several studies have correlated the timing and duration of embryo developmental events to implantation and pregnancy, and have proposed an algorithm to select an embryo for transfer with presumably the highest implantation potential (8–13). The application of such time-lapse algorithms has therefore been proposed as a method to improve embryo selection, and this has been reported to increase clinical pregnancy rates (8, 10, 14).

However, the generalizability of these algorithms still remains questionable, as efforts that have been made to externally validate them in

Received May 10, 2017; revised October 1, 2017; accepted October 27, 2017.
A.S. has nothing to disclose. C.V. has nothing to disclose. S.C. has nothing to disclose. S.K. has nothing to disclose. W.L. has nothing to disclose.
Reprint requests: Ashleigh Storr, Pg.D., IVF Australia Eastern Suburbs, Maroubra, 1/225 Maroubra Road, Maroubra, NSW 2035, Australia (E-mail: ashleigh.storr@outlook.com).

Fertility and Sterility® Vol. ■, No. ■, ■ 2017 0015-0282/$36.00

https://doi.org/10.1016/j.fertnstert.2017.10.036

different laboratories have not produced encouraging results (15, 16). For this reason, doubts have been expressed as to whether these algorithms are clinically applicable in different settings, or whether they may be site specific (16–19). Ideally, the external validity of any proposed time-lapse algorithm should be tested in the context of a randomized controlled trial (RCT). However, given that RCTs are costly and demanding in resources, it might be useful to first assess the preclinical validity of these algorithms. This would allow for a better selection of interventions that should be given priority to be tested in an RCT setting, and hence optimize the use of the frequently limited research resources in the in vitro fertilization (IVF) laboratory.

If the time-lapse algorithms were indeed able to identify the embryo with the presumably highest implantation potential, then it would be expected that there would be significant agreement between them. If this is not the case, then this highlights the discrepancies between algorithms and might reflect their limitations when applied in a different clinical setting. Furthermore, it would be interesting to assess the agreement of these algorithms with embryologists in selecting the embryo with presumably the highest implantation potential (the embryo that is usually selected first for transfer). If this agreement is high, then the expected benefit in terms of pregnancy rates is expected to be small and this should be considered when designing a relevant RCT. If this agreement is not high, then such an algorithm might have the potential to increase pregnancy rates substantially, and, for some researchers, such an algorithm might be prioritized for further evaluation in an RCT. For this reason, the aim of the current study was to evaluate the agreement between published time-lapse algorithms as well as the agreement between these algorithms and embryologists in selecting an embryo for transfer under a single day-5 transfer policy.

MATERIALS AND METHODS

Study Design

This prospective study, approved by IVFAustralia's Research and Development Committee (Project 088/March 4, 2013), compared the agreement between published time-lapse algorithms in selecting the single best day-5 embryo for transfer (i.e., the embryo with presumably the highest implantation potential) in 100 different cycles. The agreement between these algorithms and 10 embryologists working at five different clinics within IVFAustralia in New South Wales was also assessed.

Survey Design and Study Participants

Cycles for this study (n = 100) were chosen from a database of EmbryoViewer (Vitrolife) images collected from previously performed cycles from a single center where the EmbryoScope (Vitrolife) is available (7). An independent embryologist who was not included in the panel of 10 participating embryologists included the cycles based on a single criterion: the presence of at least two embryos on day 5 that could be considered for transfer (i.e., at least two embryos that had not arrested), which resulted in a total of 100 cycles including 428 embryos analyzed.

The same independent embryologist annotated each embryo, and a single two-dimensional (2D) day-5 image of 340 × 340 mm [72 pixels/inch] (EmbryoViewer; Vitrolife), taken before the time of transfer, was used in a questionnaire that was developed with a Web survey designer (SurveyMonkey Inc.) (Supplemental Fig. 1, available online). In this survey, the participating embryologists were asked which embryo they would select for transfer within each cohort, using conventional morphologic criteria based on the day-5 images they were shown, as previously described elsewhere (7). The included cycles were from a preclinical phase where the EmbryoScope was used as a standard incubator, and the embryologists did not apply any algorithm during embryo selection for transfer. The participants were asked to select only the best day-5 embryo for transfer, just as they would if they were in a laboratory setting, and they were allowed to complete the survey in more than one sitting (7).

Algorithm Selection

A PubMed search was performed on December 15, 2016, to identify publications that evaluate algorithms aiming to improve implantation or pregnancy rates. The search was performed using the following keywords: time-lapse, embryo, algorithm, implantation, and pregnancy. Supplemental Table 1 (available online) provides a summary of the publications identified (8–13, 18, 20–31) and the reasons some of them may have been excluded from the current study.

The suitable algorithms that were identified (8–13) were then applied to the cohort of cycles in the current study according to the methods outlined in each publication. The algorithms described in Meseguer et al. (9) (algorithm A) and Basile et al. (8) (algorithm D) are based on implantation data; they select the best embryo for transfer by morphologically screening embryos and then assessing them for the presence of exclusion criteria. The two algorithms then follow different hierarchical classification trees with eight morphokinetic scoring levels (A+ as the highest to F as the lowest).

The Conaghan et al. (28) algorithm (algorithm B) is based on early cleavage time intervals and includes two categories: Eeva high and Eeva low. The VerMilyea et al. (10) algorithm (algorithm C) is an extension of algorithm B, and it includes three categories: Eeva high, Eeva medium, and Eeva low.

The algorithm described in Goodman et al. (13) (algorithm E) is specifically designed for day-5 embryo selection and is based on previously established morphokinetic parameters. This algorithm initially screens embryos by means of morphology; the embryo with the highest morphokinetic score is selected based on positive and negative points (maximum of 4 and minimum of −2).

The algorithm described in Liu et al. (11) (algorithm F) is based on Known Implantation Data (KID). This algorithm first screens embryos based on morphology and abnormal cleavage patterns; then it generates a score derived from a decision tree using a morphokinetic algorithm with 5 scoring levels (A+ as the highest to F as the lowest). The KIDScore algorithm