Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review

Sir,

We have read the recent article by Kaser and Racowsky (2014) with great interest and we highly welcome this thorough systematic review. We entirely agree with the main conclusions as presented in the abstract. There are currently no high-quality data to firmly support the clinical use of this technology for selection of preimplantation embryos. Prospective studies are needed to clarify the role.

We would, however, like to discuss the results from the review upon which Drs Kaser and Racowsky base this conclusion.

The authors thoroughly review the time-lapse studies that have presented data on clinical outcome and present the results of their descriptive comparison, including their interpretation of the results from our prospective study (Kirkegaard et al., 2013).

Amongst others the authors use the results from our study to argue that there are no differences in timing between the pregnant and the non-pregnant group of all the measured parameters. Our objection is that our study was not powered to test pregnancy as a clinical outcome for all the parameters that the authors review. Accordingly, we specifically desisted from drawing any conclusions on the ability of time-lapse parameters to predict pregnancy in general. This is clearly stated in our paper. Following standard scientific conduct, we did publish timings of all the parameters in the pregnant and non-pregnant group, yet acknowledged that the study was powered only to test the parameters from the targeted logistic regression analysis. We even clarified this in a response to a letter addressing the sample size (Kirkegaard et al., 2014). We believe that the underpowered sample size entails a high risk of falsely concluding that there is no difference. For example, we did not test appearance, abuttal, syngamy, and breakdown of the male and female pronucleus (PN) as predictors of implantation in our logistic regression analysis (as stated in the review), but only PN breakdown. Therefore it is hardly justified to state that we did not find any difference in the above parameters, without acknowledging the lack of power to detect such differences. This is true for several of the conclusions the authors draw from our publication, including the conclusion that we found no difference between implanting and non-implanting embryos in terms of cleavage and blastocyst kinetics in general. We consider it plausible that the majority of the other published studies are far too small to detect any presumed differences in timing with regard to pregnancy. No randomized controlled studies of single embryo transfers have been published so far. We therefore find it very poorly supported, that reliable prediction of blastocyst formation may be the main advantage of TLM, as stated in the review.

It is correct that we conclude that TLM may decrease variability (Sundvall et al., 2013). But as the cited study involved manual, in contrast to computer-assisted annotation, it cannot be stated that the reduced variability is a result of the semi-quantification. The statement that TLM may decrease intra- and inter-observer variability among embryologists, as a result (our underlining) of computer-assisted annotation of developmental milestones and semi-quantitative process for embryo evaluation, is therefore unsupported.

In summary, we entirely agree that larger prospective studies with clinical outcomes are needed to clarify the role of time-lapse. That the existing literature suggests no association with implantation potential is in our opinion so far unjustified due to several factors, most importantly due to lack of power of the studies. This was acknowledged in the original publications, but unfortunately not in the review.

References


Reply: Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review

Dear Sir,

We thank Drs Kirkegaard and Ingerslev for their interest in our review and for their pioneering work in the field of time-lapse monitoring (TLM). Several points they raise in their Letter to the Editor deserve attention.

First, we would like to clarify that we do not maintain, as they state, that ‘there are no differences in timing between the pregnant and the non-pregnant group of all the measured parameters’ for the included studies—a conclusion that in fact was drawn in the Abstract of the cited Kirkegaard et al. (2013) publication, albeit with their acknowledged limitations in power. Rather, in our collated data we note that some parameters have been shown to be predictive of implantation (e.g. time to pronuclear breakdown, duration of the 2-cell and 3-cell stages, etc.).

We then proceed to argue that our study and in particular the results from the review upon which Drs Kaser and Racowsky base their conclusions is essentially flawed due to lack of power. However, we then state: ‘Following standard scientific conduct, we did publish timings of all the parameters in the pregnant and non-pregnant group, yet acknowledged that the study was powered only to test the parameters from the targeted logistic regression analysis. We even clarified this in a response to a letter addressing the sample size (Kirkegaard et al., 2014).’

This is not true. In our reply, we did indeed state that ‘there are no differences in timing between the pregnant and the non-pregnant group of all the measured parameters’ for the included studies—a conclusion that in fact was drawn in the Abstract of the cited Kirkegaard et al. (2013) publication, albeit with their acknowledged limitations in power. Rather, in our collated data we note that some parameters have been shown to be predictive of implantation (e.g. time to pronuclear breakdown, duration of the 2-cell and 3-cell stages, etc.).

In summary, we entirely agree with the main conclusions as presented in the abstract. There are currently no high-quality data to firmly support the clinical use of this technology for selection of preimplantation embryos. Prospective studies are needed to clarify the role.

We thank Drs Kirkegaard and Ingerslev for their interest in our review and for their pioneering work in the field of time-lapse monitoring (TLM). Several points they raise in their Letter to the Editor deserve attention.

First, we would like to clarify that we do not maintain, as they state, that ‘there are no differences in timing between the pregnant and the non-pregnant group of all the measured parameters’ for the included studies—a conclusion that in fact was drawn in the Abstract of the cited Kirkegaard et al. (2013) publication, albeit with their acknowledged limitations in power. Rather, in our collated data we note that some parameters have been shown to be predictive of implantation (e.g. time to pronuclear breakdown, duration of the 2-cell and 3-cell stages, etc.).

In summary, we entirely agree with the main conclusions as presented in the abstract. There are currently no high-quality data to firmly support the clinical use of this technology for selection of preimplantation embryos. Prospective studies are needed to clarify the role.

We thank Drs Kirkegaard and Ingerslev for their interest in our review and for their pioneering work in the field of time-lapse monitoring (TLM). Several points they raise in their Letter to the Editor deserve attention.

First, we would like to clarify that we do not maintain, as they state, that ‘there are no differences in timing between the pregnant and the non-pregnant group of all the measured parameters’ for the included studies—a conclusion that in fact was drawn in the Abstract of the cited Kirkegaard et al. (2013) publication, albeit with their acknowledged limitations in power. Rather, in our collated data we note that some parameters have been shown to be predictive of implantation (e.g. time to pronuclear breakdown, duration of the 2-cell and 3-cell stages, etc.).
and times to the 5-cell, 8-cell and various blastocyst stages), while others have not. It happens that none of the parameters assessed in the Kirkegaard et al. (2013) analysis were significantly different between embryos that implanted and those that failed to implant, perhaps due in part to type II error. As undertaken by Kirkegaard and her colleagues (Kirkegaard et al., 2013; Supplementary Table 1), we also systematically considered each parameter and reported whether or not various groups found that the mean or median values differed significantly between implanted and non-implanted embryos. We emphasized that no single parameter has consistently been shown to correlate with clinical outcome, likely due to small sample sizes and study heterogeneities (type of media, 5–6% versus atmospheric oxygen tension, single versus group culture, mode of fertilization, and day of embryo transfer). As a result of these heterogeneities, and the inconsistencies in time-lapse nomenclature (Table II in Kaser and Racowsky, 2014), summary measures such as risk ratios or differences in means could not be calculated.

We have therefore proposed a standardized nomenclature and syntax (Table IV) to allow users to describe any milestone along the preimplantation developmental timeline, whether using brightfield or darkfield illumination. Adoption of such a nomenclature will permit meaningful comparison of time-lapse parameters among future studies so that the problem of small sample sizes and limited power in any one particular study, as pointed out by Drs Kirkegaard and Ingerslev, might be overcome with meta-analysis.

Second, we reaffirm our viewpoint that available evidence suggests that the main advantage of TLM is blastocyst prediction. Because current studies are underpowered, we believe that use of time-lapse measures for prediction of clinical outcomes is premature. As further evidence accumulates, it may become apparent that there are reliable markers for implantation.

Third, it was not our intent to imply that the improved intra- and inter-observer variability among embryologists is a result solely of computer-assisted annotation; rather, we believe that the semi-quantitative process of annotation itself, not the mode of the annotation, is what improves the kappa statistic of conventional morphology. This is evidenced by an observed reduction in variability with both automated cell-tracking combined with conventional morphology (Conaghan et al., 2013) and manual annotation (Sundvall et al., 2013).

Fourth, contrary to the claim by Drs Kirkegaard and Ingerslev, we specifically drew attention to the small sample sizes of many included studies as a key limitation. We thank these authors for further underscoring this important point about their own study, as well as that of several of the other studies included in our review. We acknowledge that our conclusions may be affected by the lack of appropriately powered studies and reiterate that ‘our recommendations for the adoption of this technique are thus limited by the available literature.’

Finally, and perhaps most importantly, we agree with Drs Kirkegaard and Ingerslev that further prospective studies (with, indeed, adequate power to detect the outcome of interest) are warranted. Our hope is that our review highlights knowledge gaps in this exciting field and provides a standardized nomenclature and syntax to allow meaningful comparisons among future studies, regardless of the type of TLM system used to acquire the images.

Conflict of interest

C.R. acts as a consultant (scientific advisor) for Auxogyn, Inc.

References


Daniel J. Kaser and Catherine Racowsky*Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
*Correspondence address. cracowsky@partners.org
doi:10.1093/humupd/dmu045
Advanced Access publication on July 17, 2014