Accepted Manuscript

Title: Limitations of a time-lapse blastocyst prediction model: a large multicentre outcome analysis

Author: Kirstine Kirkegaard, Alison Campbell, Inge Agerholm, Ursula Bentin-Ley, Anette Gabrielsen, John Kirk, Shabana Sayed, Hans Jakob Ingerslev

PII: S1472-6483(14)00246-6
DOI: http://dx.doi.org/doi:10.1016/j.rbmo.2014.04.011
Reference: RBMO 1147

To appear in: *Reproductive BioMedicine Online*

Received date: 10-9-2013
Revised date: 10-2-2014
Accepted date: 23-4-2014

Please cite this article as: Kirstine Kirkegaard, Alison Campbell, Inge Agerholm, Ursula Bentin-Ley, Anette Gabrielsen, John Kirk, Shabana Sayed, Hans Jakob Ingerslev, Limitations of a time-lapse blastocyst prediction model: a large multicentre outcome analysis, *Reproductive BioMedicine Online* (2014), http://dx.doi.org/doi:10.1016/j.rbmo.2014.04.011.

Testing an blastocyst prediction model on clinical outcome

Limitations of a time-lapse blastocyst prediction model: a large multicentre outcome analysis

Kirstine Kirkegaard[a, b] (corresponding author), Alison Campbell[h], Inge Agerholm[c], Ursula Bentin-Ley[d], Anette Gabrielsen[e], John Kirk[f], Shabana Sayed[g], and Hans Jakob Ingerslev[a,b]
a: Fertility Clinic, Aarhus University Hospital, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark
b: Health, Aarhus University, Vennelyst Boulevard 9, 8000 Aarhus C, Denmark
c: The Fertility Clinic, Braedstrup Hospital, Sygehusvej 20, 8740 Brædstrup, Denmark
d: Dansk Fertilitetsklinik, Seedorffs Vaenge 2, 2000 Frederiksberg, Denmark
e: Ciconia Aarhus Privathospital, Saralyst Allé 50, 8270 Højbjerg, Denmark
f: Maigaard Fertility Clinic, Jens Baggesensvej 88 H, 8200 Aarhus N, Denmark
g: Klinikk Hausken, Karmsundgata 59, 5531 Haugesund, Norway
h: CARE Fertility Group, John Webster House, 6 Lawrence Drive, Nottingham Business Park, Nottingham NG8 6PZ, United Kingdom
Contact information (email) and academic degrees:
Kirstine Kirkegaard (corresponding author): MD, PhD
kirstine.kirkegaard@ki.au.dk, +4520917620. Fertility Clinic, Aarhus University Hospital, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark
Alison Campbell: BSc (hons), M. Med. Sci.
Alison.Campbell@carefertility.com
Inge Agerholm: MSc, PhD inge.e.agerholm@horsens.rm.dk
Ursula Bentin-Ley: MD, PhD ursula@danfert.dk
Anette Gabrielsen: MSc ag@ciconia.dk
Shabana Sayed: MSc. (Biotech), MCE(Embryology)
shabana@klinikkhausken.no
John Kirk:MD john.kirk@dadlnet.dk
Hans Jakob Ingerslev:MD, DrMedSci jakob.ingerslev@skejby.rm.dk

Abstract

The goal of embryo selection models is to select embryos with the highest reproductive potential, whilst minimizing the rejection of viable embryos. Ultimately, any embryo selection model must be tested on clinical outcome. We therefore retrospectively tested a published blastocyst prediction model on a large combined set of transferred embryos with known clinical outcome. The model was somewhat effective in that it predicted a relative increase of 30% for implantation in the model-selected group of embryos, but the model's results did not match observations in the test cohort because it rejected a large proportion of embryos from the test cohort that actually resulted in pregnancy. This

Comment [KMT1]: Author: please confirm edit or amend.

hypothetical experiment highlights the limitations of predicting blastulation only. Crucially, it illustrates that both sensitivity and specificity are important parameters when developing embryo selection models for prospective clinical use.

Keywords: *assisted reproduction, embryo selection, prediction model, time lapse*

Introduction

Time-lapse imaging of human preimplantation embryos has become rapidly integrated in IVF laboratories. The proposed advantages, such as uninterrupted embryo culture, flexibility in timing, improvement of documentation procedures, quality control and management and, in particular, the introduction of dynamic markers of embryo quality, have altogether stimulated a profound interest in time-lapse technology. While a large number of publications consolidate that timing of development differs between viable and nonviable embryos (Herrero and Meseguer, 2013), only a few publications offer clinically applicable models of embryo selection (Conaghan *et al.*, 2013, Meseguer *et al.*, 2011, Campbell *et al.*, 2013). Yet, as recently demonstrated, a proposed multivariate hierarchical selection model was not transferable from one clinical setting to another without modification (Best *et al.*, 2013). It has been speculated that a less-complex model, such as the one recently developed and applied by Conaghan *et al.* (2013) that categorized embryos into groups with either high or low likelihood of forming 'usable blastocysts', could be applicable to other clinics. The model has, however, not been evaluated with regard to clinical outcome. To test this hypothesis and correlation between the published time intervals and clinical outcome, we retrospectively applied the same model to a large set of transferred embryos from independent clinics.

Retrospective testing of a blastocyst prediction model

Seven clinics from three different countries participated by contributing data on clinical outcome following embryo transfer (fetal heart beat) and timing of cellular divisions until day 3, obtained using time-lapse monitoring (EmbryoScope, FertiliTech, Denmark). The first division was annotated t_2, second division t_3 and the third division t_4.

A total of 1519 transferred embryos with known outcome for implantation from cycles with single (n = 517) or double (n = 501) embryo transfer were included. In order to be able to relate each embryo's fate after transfer with its individual morphokinetic profile, only cycles with two or no fetal heart beats were included where double-embryo transfers were performed. This implies that the presented pregnancy rates are lower and not directly comparable with treatment success rates. Patient stimulation, IVF/intracytoplasmic sperm injection and embryo culture was performed according to standard procedures at each site. Embryos were graded and selected according to each clinic's routine methodology and one or two embryos were selected for transfer. Embryo transfer was predominantly performed on day 2 or day 3.

In order to test whether time intervals published in Conaghan *et al.* (2013) correlated with clinical outcome, we retrospectively grouped the transferred embryos into usable and nonusable embryos based on the model's values for these time-lapse intervals. The model was found to predict a high chance of usable blastocyst formation (defined as a blastocyst suitable for either transfer or freezing) if time between first and second cytokinesis ($t_3 - t_2$) was 9.33–11.45 h and time between second and third cytokinesis ($t_4 - t_3$) was 0–1.73 h. Likewise, embryos were predicted to have a low chance of forming usable blastocysts if $t_3 - t_2$ and $t_4 - t_3$ were longer than these time intervals. We calculated the relative difference in implantation (%) between the usable group and the entire cohort, odds ratio for implantation in the usable compared with the nonusable group and the percentage of nonusable embryos that resulted in implantation. Data were used to generate a receiver operating characteristic curve and to calculate area under the curve for implantation.

Comment [KMT2]: Author: please confirm edit or amend, according to table note.

Data for implantation are presented in **Table 1**. The relative difference in implantation rate between the entire cohort and the embryos categorized as usable by the test model was 30.0%. The odds ratio for implantation between usable and nonusable was 1.60. The sensitivity was 0.38 and the specificity was 0.85. Furthermore, 50.6% of the embryos that resulted in pregnancy were categorized as nonusable according to the model. The area under the curve for prediction of implantation was 0.57 (**Figure 1**).

In other words, applying the test model retrospectively to transferred embryos from the independent clinics would have provided an increase of 30.0% in implantation rate for embryos grouped as usable compared with the entire test cohort. Notably, out of the embryos that actually implanted, 50.6% were categorized as having low chance of being usable. This indicates that relying on such a model would bring a substantial risk of deeming viable embryos nonusable.

The ultimate goal of embryo selection models is to positively select embryos with the highest reproductive potential, notably without rejecting viable embryos. The premise of this hypothetical experiment is that the test model predicts the formation of usable blastocysts (i.e. blastocysts to be either transferred or frozen. If embryos are selected for transfer/freezing on day 3 (as suggested by the authors), it ultimately follows that blastocysts with a low chance of being usable are to be discarded if the model is applied uncritically. Principally, embryos from this study population that have implanted would have been discarded with day-3 transfer and application of the time-lapse-based selection model.

This hypothetical experiment illustrates the risks of defining too narrow time intervals for optimal division in order to achieve a high specificity at the expense of a low sensitivity. It thus underlines the importance of carefully considering that a model must not only provide a substantial increase in implantation but also, equally important, that a low rejection rate of viable embryos is secured. This very important point is demonstrated by applying the time-lapse criteria on an unprecedented large set of transferred embryos with a known outcome.

From blastocyst to pregnancy and beyond

The test model was developed in order to identify viable embryos from a cohort, while the study population is constituted by embryos selected for transfer with clinical outcome as the endpoint. Thus, both the endpoint and the study population differ between the two studies. It would be expected that a model that predicts blastocyst development would be different from a model that predicts clinical outcome. In our opinion, it would, however, be expected that a model that predicts blastocyst development would positively select more embryos than a selection model predictive of implantation and clinical pregnancy, as embryos resulting in pregnancy would constitute a subgroup of embryos that develop into blastocysts. We do not, however, find any explanation, neither in the different study populations nor in the different endpoints, as to why a large proportion of embryos which were rendered 'unusable' by the model as they fell outside the model selection criteria resulted in implantation. In our opinion, the most likely explanation is the narrow time intervals for optimal cellular division.

Ultimately, any embryo selection model must be tested on clinical outcome, as blastocyst development is a surrogate endpoint, preferably in a prospective study. We believe that the approach of retrospectively testing the criteria on transferred embryos with known outcome is justified in this case, as it aids the design of future prospective studies.

Our study supports an approach where models are developed with appropriate concern for low rejection of viable embryos, where clinical outcome is used as an endpoint and where the model is individually adjusted to specific settings and validated prior to implementation.

References

CAMPBELL, A., FISHEL, S., BOWMAN, N., DUFFY, S., SEDLER, M. and THORNTON, S. 2013. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reprod Biomed Online* 27, 140–146

CONAGHAN, J., CHEN, A. A., WILLMAN, S. P., IVANI, K., CHENETTE, P. E., BOOSTANFAR, R., BAKER, V. L., ADAMSON, G. D., ABUSIEF, M. E., GVAKHARIA, M., LOEWKE, K. E. and SHEN, S. 2013. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril,* 100, 412–419 e5.

HERRERO, J. and MESEGUER, M. 2013. Selection of high potential embryos using time-lapse imaging: the era of morphokinetics. *Fertil Steril,* 99, 1030–4.

MESEGUER, M., HERRERO, J., TEJERA, A., HILLIGSOE, K. M., RAMSING, N. B. and REMOHI, J. 2011. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod,* 26, 2658–71.

BEST, L., CAMPBELL, A., DUFFY, S., MONTGOMERY, S., FISHEL, S. 2013. Session 57: Does one model fit all? Testing a published embryo selection algorithm on independent time-lapse data. *Human Reproduction,* 28, i87-i90

Declaration: The authors report no financial or commercial conflicts of interest.

Table 1. Implantation data for embryos categorized as usable or nonusable according to the test model.

	Implanted	*Not implanted*	*Implantation rate*
Usable	131	445	22.7
Nonusable	134	809	14.2
Entire cohort	265	1254	17.4

Values are *n* or %.

Usable: $t_3 - t_2$ = 9.33–11.45 h and $t_4 - t_3$ = 0–1.73 h. Unusable: $t_3 - t_2$ more than 9.33–11.45 h and $t_4 - t_3$ more than 0–1.73 h.

Figure 1. Receiver operating characteristic curve for prediction of pregnancy by the parameters $t_3 - t_2$ and $t_4 - t_3$ from the test model.