Reproductive BioMedicine Online (2015) 31, 459–471

www.sciencedirect.com
www.rbmonline.com

SYMPOSIUM: QUALITY MANAGEMENT IN ASSISTED REPRODUCTIVE TECHNOLOGY REVIEW

Quality control and standardization of embryo morphology scoring and viability markers

Kersti Lundin *, Aisling Ahlström

Reproductive Medicine, Sahlgrenska University Hospital, Göteborg 413 45, Sweden
* Corresponding author. *E-mail address:* kersti.lundin@vgregion.se (K Lundin).

Kersti Lundin obtained her PhD in Zoophysiology at the University of Gothenburg, Sweden in 1991. She thereafter started as an embryologist and researcher at the Unit of Reproductive Medicine, Sahlgrenska Hospital, Gothenburg, where she became Laboratory Director of Reproductive Medicine in 1997 and Associate Professor in 2004. Her main subjects of interest are basic and clinical embryology, including embryo development and cryopreservation.

Abstract A so-called 'good-quality embryo' may be defined as an embryo that has the potential to implant into the uterine endometrium and give rise to the birth of a healthy child. A standardized and objective scoring of embryo 'quality' is therefore crucial in the classification and selection of embryos. However, embryo scoring is still being performed mainly via ocular evaluation, which often results in different interpretations of embryo quality. The addition of viability markers, such as measuring gene expression or the uptake/release of metabolites, proteins or RNA/DNA molecules in the culture media, would increase the possibility of standardized measurements. However, no single biomarker has yet been introduced into standard clinical practice, mainly due to the complexity of the techniques and the influence of biological variations and differences in culture conditions. In this paper different methods for the scoring of embryos and the possibility of standardizing and implementing quality control systems are discussed.

KEYWORDS: embryo, human, IVF, quality control, validation

Introduction

Although the efficiency of an IVF laboratory can to a large extent be controlled by rules and regulations, such as the International Organization for Standardization (ISO) norms, a majority of the work is still dependent on subjective evaluations and decisions. In particular, morphological assessment and scoring of embryos is dependent upon a number of variables, being more or less subjective.

The subjective nature of morphological scoring by an embryologist, and in addition the existence of confounding (laboratory) factors such as differences in culture media and culture environment, as well as different handling of oocytes and embryos in the laboratory, makes it difficult to compare

http://dx.doi.org/10.1016/j.rbmo.2015.06.026

embryo scores – and thereby also success – rates between clinics.

The aim of this paper is to discuss the possibility of standardizing and implementing quality control systems for the scoring of embryos and to evaluate alternative methods being currently implemented into standard clinical practice.

Morphological variables predictive of implantation and live birth

Numerous embryo morphology scoring systems have been developed throughout the years, and there are several classification schemes being used. Some list the embryo features individually, often ranked according to their supposed importance for embryo viability, while others use an algorithm to calculate a cumulative, single score based on a number of (weighted) features. Main morphological features scored include the number of cells, grade of fragmentation, cell size and multinucleation/number of mononucleated cells; and for blastocyst stage embryos, expansion grade and the status of the inner cell mass (ICM) and the trophectoderm.

Ideally, embryo scoring should be included in quality assurance schemes, which should be fast and simple and based on variables with proven predictive power for live birth. Van Loendersloot et al. (2014) constructed a ranking model for day 3 transfers, using one data set for constructing the model and another data set for validation of the method (Van Loendersloot et al., 2014). They found that blastomere numbers on day 2 and day 3, morphological score on day 3 and morula on day 3 (yes/no) were correlated with implantation. These results were very similar to those shown by Van Royen et al. (2001), who in short demonstrated that the presence of four cells on day 2 and eight cells on day 3, together with a low fragmentation grade on day 3, results in the best implantation rates (Van Royen et al., 2001). It is important to note, however, that in the study by Loendersloot et al., the time range for the scoring could be up to 5 h, a source of variation that will be discussed later in this review.

In two other papers, regression analyses of large data sets have been used to standardize embryo scoring and build predictive models for day 2 (Holte et al., 2007) and for day 2/day 3 (Racowsky et al., 2009). In both these models, it was found that blastomere number was the most powerful predictor for implantation, while in Racowsky's model, the importance of fragmentation for day 3 increased compared with day 2.

In a prospective study of 6252 single embryo transfer cycles where the model for day 2 embryos by Holte et al. (2007) was used, Rhenman et al. (2015) demonstrated that the variables blastomere number, proportion of mononucleate blastomeres and degree of fragmentation were all predictive for live birth (Rhenman et al., 2015). The previously integrated variable, equal sized blastomeres, was correlated with live birth in the univariate analysis, but did not come out as predictive in the regression analysis.

For blastocysts, grading three morphological features, degree of blastocoele expansion and hatching stage, ICM, and trophectoderm, have all been shown to be correlated with pregnancy and live birth. Furthermore, scoring systems taking into account the appearance of all three features have been proven to significantly improve selection of viable blastocysts and prediction of clinical outcome compared with scoring of a single feature (Balaban et al., 2006; Dokras et al., 1991, 1993; Gardner and Schoolcraft, 1999; Richter et al., 2001; Shapiro et al., 2008b). The most prominent grading system described by Gardner and Schoolcraft (1999) has been validated by several studies to show that the transfer of two top-scoring blastocysts with high grades for all three features achieves the highest implantation rates (Balaban et al., 2000, 2006; Gardner et al., 2000, 2004). However, retrospective studies attempting to determine the independent predictive strength of each feature and even rank their importance are so far inconclusive. In a multicentre trial by Van Den Abbeel et al. (2013), expansion and hatching stage was determined as the only significant predictor of live birth ($P = 0.002$), while Ahlstrom et al. (2011a) and Hill et al. (2013) found trophectoderm grade to be the only statistically significant independent predictor of live birth (Ahlstrom et al., 2011a; Hill et al., 2013; Van Den Abbeel et al., 2013).

Standardization and proficiency

When discussing and implementing embryo scoring and embryo quality, it is important to consider the potential variations in embryo scoring within each individual (intra-observer), between individuals (interindividual) and between centres (intercentre). These – often considerable – variations will influence the interpretation of embryo quality, and thereby the data being used for analysis of correlations with implantation and live birth. This is a problem, both during interpretation of published studies regarding impact on success rates and during clinical application of criteria used to select embryos.

Intra- and interobserver analyses of embryo scoring may have different goals. Firstly, to analyse the agreement when scoring individual features of the embryo such as number of cells, grade of fragmentation, etc. Secondly, to determine the agreement in classifying the embryos on a scale ('top', 'good', 'moderate', 'poor') and thirdly, to select the embryo(s) most suitable for transfer and cryopreservation. Thus, the ability to transform the individual scoring to an overall classification (top, good, fair, poor) and thereafter to rank it for a clinical decision within each cohort of embryos is just as important as being able to correctly count the number of cells and the percentage of fragments.

In a paper by Baxter Bendus et al. (2006), inter- and intra-observer variations for 26 embryologists who graded day 3 embryos in video sessions were compared (Baxter Bendus et al., 2006). Poor interobserver agreement (median Kappa value 0.24, range 0.03–0.49) was found, while the intra-observer agreement (scoring the same embryo several times) was good (median Kappa value 0.69, range 0.44–1.00).

In another paper by Arce et al. (2006), more than 15,000 embryos were scored (with the help of an embryo atlas) by a total of 37 local IVF laboratories (Arce et al., 2006). The same embryos were then scored using 2D electronic images, by three 'central' embryologists (interobserver comparison). These embryologists had been practising embryo scoring together in joint training sessions in order to improve the interobserver agreement. It was shown that for the individual variables cell number, fragmentation and blastomere size the interobserver agreement between the three central embryologists was good to high (kappa values: 0.61–0.94),