Examining the efficacy of six published time-lapse imaging embryo selection algorithms to predict implantation to demonstrate the need for the development of specific, in-house morphokinetic selection algorithms

Amy Barrie, M.Sc.,[a] Roy Homburg, M.B.B.S.,[a] Garry McDowell, Ph.D.,[b] Jeremy Brown, Ph.D.,[c] Charles Kingsland, M.D.,[a] and Stephen Troup, Ph.D.[a]

[a] Hewitt Fertility Centre, Liverpool Women's NHS Foundation Trust, Liverpool; [b] Manchester Metropolitan University, Manchester; and [c] Edge Hill University, Ormskirk, United Kingdom

Objective: To study the efficacy of six embryo-selection algorithms (ESAs) when applied to a large, exclusive set of known implantation embryos.
Design: Retrospective, observational analysis.
Setting: Fertility treatment center.
Patient(s): Women undergoing a total of 884 in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) treatment cycles (977 embryos) between September 2014 and September 2015 with embryos cultured using G-TL (Vitrolife) at 5% O_2, 89% N_2, 6% CO_2, at 37°C in EmbryoScope instruments.
Intervention(s): None.
Main Outcome Measure(s): Efficacy of each ESA to predict implantation defined using specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic curve (AUC), and likelihood ratio (LR), with differences in implantation rates (IR) in the categories outlined by each ESA statistically analyzed (Fisher's exact and Kruskal-Wallis tests).
Result(s): When applied to an exclusive cohort of known implantation embryos, the PPVs of each ESA were 42.57%, 41.52%, 44.28%, 38.91%, 38.29%, and 40.45%. The NPVs were 62.12%, 68.26%, 71.35%, 76.19%, 61.10%, and 64.14%. The sensitivity was 16.70%, 75.33%, 72.94%, 98.67%, 51.19%, and 62.33% and the specificity was 85.83%, 33.33%, 42.33%, 2.67%, 48.17%, and 42.33%, The AUC were 0.584, 0.558, 0.573, 0.612, 0.543, and 0.629. Two of the ESAs resulted in statistically significant differences in the embryo classifications in terms of IR.
Conclusion(s): These results highlight the need for the development of in-house ESAs that are specific to the patient, treatment, and environment. These data suggest that currently available ESAs may not be clinically applicable and lose their diagnostic value when externally applied. (Fertil Steril® 2017;107:613–21. ©2016 by American Society for Reproductive Medicine.)
Key Words: Embryo development, embryo selection algorithm, morphokinetics

Discuss: You can discuss this article with its authors and with other ASRM members at https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/13000-22789

Received July 18, 2016; revised November 4, 2016; accepted November 7, 2016; published online January 6, 2017.
A.B. has nothing to disclose. R.H. has nothing to disclose. G.M. has nothing to disclose. J.B. has nothing to disclose. C.K. has nothing to disclose. S.T. has nothing to disclose.
Reprint requests: Amy Barrie, M.Sc., The Hewitt Fertility Centre, Liverpool Women's NHS Foundation Trust, Crown Street, Liverpool L8 7SS, United Kingdom (E-mail: amy.barrie@lwh.nhs.uk).

Fertility and Sterility® Vol. 107, No. 3, March 2017 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2016.11.014

Traditional methods for embryo selection have been used for over 20 years. Numerous morphologic parameters are thought to be useful for correct embryo selection: pronuclear morphology (Z scoring) (1, 2), polar body alignment and appearance (3, 4), appearance of cytoplasm and zona pellucida (5), early cleavage (6, 7), multinucleation (8–10), and blastomere morphology (11–13). Basic embryo grading, including the number of blastomeres, evenness in the size of the blastomeres, and the level of fragmentation, remains the gold standard for embryo selection. However, using this method in a traditional sense (with a standard bench-top incubator) has two limitations: a restricted overview of an embryo's development and the exposure of the embryo to suboptimal temperatures and gas concentrations. With the introduction of time-lapse imaging, where an image of each embryo is taken every 10 to 20 minutes, more intricate embryo parameters can be viewed while leaving the embryos in an undisturbed environment. As the availability of time-lapse technologies increased, attention was first focused on assessing their clinical safety. Once the safety had been established and the available technologies were validated for clinical use (14–18), research then turned to determining how the time-lapse imaging systems could be used to increase pregnancy rates through in-depth embryo analysis and an undisturbed culture system.

Through the research performed previously and subsequently, many morphokinetic parameters were identified that correlated with the embryo's ability to create a pregnancy both in humans and animals: the appearance and disappearance of pronuclei and nuclei at each cell stage (3, 19–21), the length of time between early cytokinesis (22–29) and initiation of blastulation (30). Further embryologic phenomena have been observed using time-lapse imaging, including the reabsorption of fragments (31), direct cleavage of cells within embryos from one to three cells (32), and reverse cleavage (33). These phenomena have been shown to affect an embryo's implantation potential to varying degrees, but their discovery could lead to more effective embryo selection within a laboratory using time-lapse technology.

Single-embryo parameters such as those mentioned here have been linked to embryo viability (18), and now these parameters have been used to develop embryo-selection algorithms (ESAs). These ESAs seek to combine a number of morphokinetic parameters that have been linked to an embryo's viability expressed as the formation of a blastocyst, implantation, or a live birth. Here, the efficacy of six published ESAs for predicting an embryo's viability was examined, expressed as implantation rate (IR), in a clinically applicable setting (21, 27, 30, 34, 35) to demonstrate the need to develop specific, in-house ESAs. The ESAs examined were selected based on their clinical applicability to the test site, assessed superficially before analysis.

MATERIALS AND METHODS

This investigation was a single site, retrospective observational design approved by the North West Research Ethics Committee (ref: 14/NW/1043) and the institutional review board where necessary. All procedures and protocols complied with United Kingdom regulations (Human Fertilisation and Embryology Act, 1990, 2008). The data were obtained from 884 treatment cycles between September 2014 and December 2015. Clinical pregnancy was confirmed by the presence of a fetal heartbeat at ultrasound scan at 6 weeks' gestation. All treatments included in this analysis were from known implantation embryos; a single-embryo transfer or a double-embryo transfer where the transfer of two embryos resulted in either a negative test or two fetal heartbeats.

Ovarian Stimulation

Pituitary down-regulation was achieved using either a gonadotropin-releasing hormone agonist (buserelin, Suprecur; Sanofi Aventis) or antagonist (cetrorelix acetate, Cetrotide; Merck Serono). Ovarian stimulation was performed using urine-derived or recombinant follicle-stimulating hormone (Progynova [Bayer Germany]; Fostimon and Merional [IBSA]; Menopur [Ferring Fertility]; or Gonal f [Merck Serono]). Doses were adjusted based on the patient's demographic and response. Patients were given 5,000 IU of subcutaneous human chorionic gonadotropin (Gonasi HP; IBSA Pharmaceuticals) 36 hours before oocyte collection. Luteal support was provided via 400 mg of progesterone pessaries, twice daily (Cyclogest; Actavis), until the pregnancy test was taken.

Oocyte Retrieval and Embryology

Ultrasound-guided oocyte collection was performed transvaginally under sedation (Diprivan; Fresenius Kabi). Collected oocyte–cumulus complexes were cultured in four-well dishes (Nunc; Thermo Scientific) with each well containing 0.65 mL of G-IVF (Vitrolife) covered with 0.35 mL of OVOIL (Vitrolife) in a standard incubator (Sanyo Multigas MCO 18M). Sperm preparation was performed using a standard gradient separation at 0.3 relative centrifugal force (rcf) for 10 minutes (ISolate; Irvine Scientific) followed by two washes at 0.6 rcf for 10 minutes using G-IVF. The oocytes destined for intracytoplasmic sperm injection (ICSI) were prepared using enzymatic (Hyase 10X; Vitrolife) and mechanical digestion. Intracytoplasmic sperm injection was performed approximately 4 hours after collection, following which all injected oocytes were placed in individual culture drops of G-TL (Vitrolife) and cultured in the EmbryoScope (Vitrolife). The oocytes destined for standard insemination had this performed approximately 4 hours after collection and replaced into a standard incubator until the fertilization check the next day. Oocytes were then checked for fertilization approximately 16 to 18 hours postinsemination (hpi), and all fertilized oocytes along with all unfertilized metaphase II oocytes were placed in individual culture drops of G-TL and cultured in the EmbryoScope.

Embryo selection was performed using the national grading scheme (36) along with an internally derived ESA. This ESA was used as an additive to morphology at the test site and only used when two or three (where double-embryo transfer was to be performed) embryos of similar morphology were available for transfer. Morphology remained the gold

standard for embryo selection. This ESA included three morphokinetic parameters: s2 (the time between three-cell [t3] and four-cell [t4]), cc3 (defined at the study site as the time between t4 and five-cell [t5]), and t5 with embryos graded in one of eight categories from A+ to D−. Embryo transfer was performed using the highest grade embryo(s) 5 days after collection. Selected embryos were cultured in EmbryoGlue (Vitrolife) for 10 to 30 minutes in a standard incubator before embryo transfer. All embryos were cultured at 37°C in 6% CO_2, 5% O_2, and 89% N_2 throughout.

Analysis of Time-lapse Information

The image interval on the EmbryoScope was set to 15 minutes with seven focal planes. Images were collected for the duration of culture immediately after ICSI or fertilization check (for IVF-derived embryos) to utilization. A single embryologist assessed images for the required morphokinetic parameters with t0 defined as the time of insemination/injection. The parameters annotated included time to pronuclear fading (tPNf), time to two-cell (t2), t3, t4, t5, eight-cell (t8), time to start of blastoculation (tSB), time to blastocyst (tB, defined when the blastocele has filled over half of the embryo and there is a <10% increase in the embryo diameter (i.e. the beginning of expansion) quantified using the line tools on the EmbryoScope instrument). From these annotations, two further annotations were calculated (s2 and cc2, the time to complete the second cell cycle). Accuracy of annotation was corroborated by the participation of the embryologist in an internal quality assurance scheme for morphokinetic analysis. Each of the ESAs (Table 1) was then retrospectively applied to the same cohort of known implantation embryos.

Statistical Analysis

Positive predictive value (PPV), negative predictive value (NPV), specificity, and sensitivity were used to determine the efficacy of each of the ESAs. These methods of measurement were chosen for the analysis because of their relationship to validity and predictive power. We defined PPV as the percentage of embryos creating a fetal heartbeat as well as a favorable ESA outcome. We defined NPV as the percentage of embryos not creating a fetal heartbeat as well as an unfavorable ESA outcome. Sensitivity was defined as the ability of the ESA to correctly classify an embryo as viable. Specificity was defined as the ability of the ESA to correctly classify an embryo as nonviable.

Each of the test measures were determined using the following calculations:

PPV = True positives/(True positives + False positives)

NPV = True negatives/(True negatives + False negatives)

Sensitivity = True positives/(True positives + False negatives)

Specificity = True negatives/(True negatives + False positives)

The likelihood ratio (LR) was determined using the following calculation:

TABLE 1

Summary of embryo selection criterion and main results.

Study	Model type	Parameter	Time frame	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC	Likelihood ratio	*P* Value
Azarello et al. 2012 (21)	Selection/deselection	PNf	>20 h 45m	16.71	85.83	42.57	62.12	0.584	1.18	.2724[a]
Basile et al. 2015 (37)	Hierarchical	t3	34–40 hpi	75.33	33.33	41.52	68.26	0.558	1.13	.006[b]
		cc2	9–12 h							
Campbell et al. 2013 (30)	Risk classification model	tSB	Low risk: tSB <96.2 hpi tB <122.9 hpi	72.94	42.33	44.28	71.35	0.579	1.26	<.0001[b]
		tB	Medium risk: tSB ≥96.2 hpi tB <122.9 hpi							
			High risk: tB ≥ 122.9 hpi							
Chamayou et al. 2013 (34)	Selection/deselection	cc3	9.7–21.0 h	98.67	2.67	38.91	76.19	0.552	1.01	.1817[a]
Cruz et al. 2012 (27)	Hierarchical	t5	48.8–56.6 h	51.19	48.17	38.29	61.10	0.517	0.99	.1402[b]
		s2	≤0.76 h							
Dal Canto et al. 2012 (35)	Selection/deselection	t8	54.9 ± 5.2 h	62.33	42.33	40.45	64.14	0.583	1.08	.1415[a]

Note: P value stated when comparing the implantation rates between categories in the applicable ESA. AUC = area under the receiver operating characteristic curve; cc2 = time between two-cell and three-cell; cc3 = time between three cell and five cell; hpi = hours post-insemination; PNf = pronuclear fading; PPV = positive predictive value; NPV = negative predictive value; s2 = time between three cell and four cell; t3 = time to three cell; t5 = time to five cell; t8 = time to eight cell; tB = time to full blastocyst; tSB = time to start of blastulation.

[a] Fisher's exact test.

[b] Kruskal-Wallis test.

Barrie. Efficacy of six published ESAs. Fertil Steril 2016.

LR = sensitivity/(1 − Specificity)

The area under the receiver operating characteristic curve (AUC) was calculated for each ESA. The IR in each category of the ESA was compared using Fisher's exact test (for ESAs with two outcome categories—true or false) and Kruskal-Wallis test (for ESAs with more than two outcome categories: A, B, C, and D). $P<.05$ was considered statistically significant. Statistical analyses were performed using the statistical package Prism 5 (GraphPad Software).

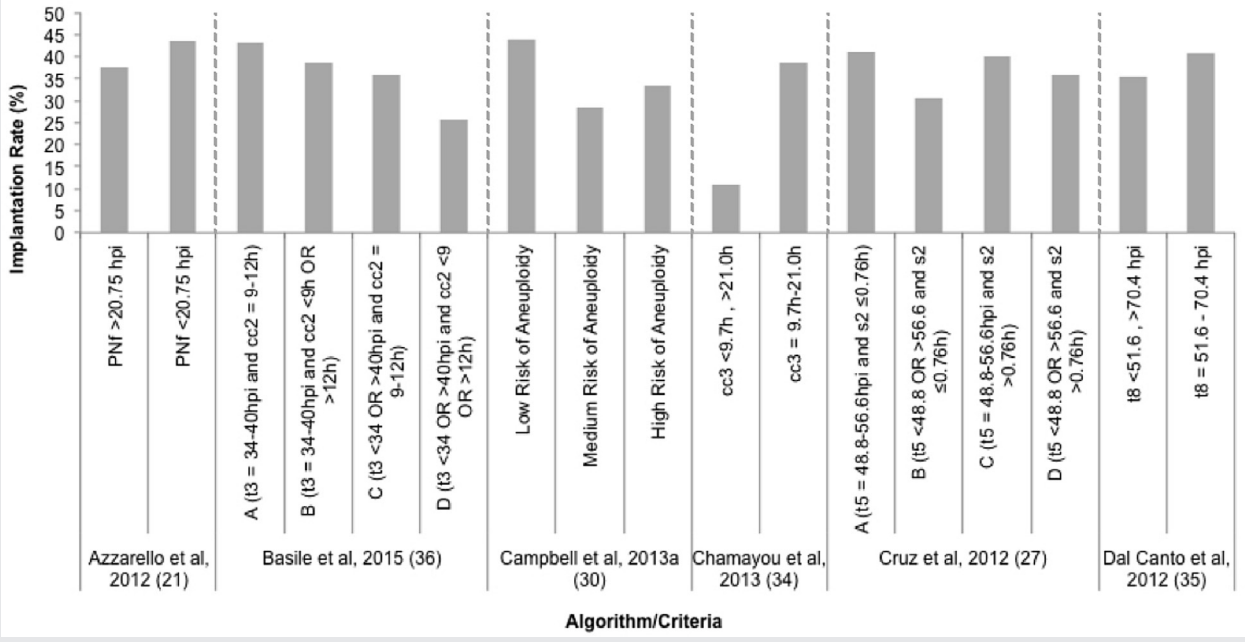
RESULTS

A total of 977 known implantation embryos from 884 treatment cycles were subject to retrospective analysis to determine the efficacy of six published ESAs (Table 1). Of these, 529 embryos were created using conventional IVF, and 448 were created using ICSI. The mean patient age was 33.44 ± 4.53 years with an average treatment attempt number of 1.37. The primary etiologies for infertility were male factor (32.2%), maternal age (4.1%), ovulatory disorders (9.9%), tubal disorders (6.6%), uterine disorders (4.1%), other (including genetic disorder) (0.2%), hormone deficiency (1%), and unexplained (41.9%). All embryo transfers were performed on day 5 (blastocyst). Ninety-three double-embryo transfers and 791 single-embryo transfers were performed. Among the treatment cycles, 50.36% were an agonist protocol; the remainder of the cycles were an antagonist protocol. An overall IR of 39.7% was achieved with 388 of the 977 embryos implanting and 589 not implanting.

The PPV for each of the ESAs did not reach above 45% in any case. The NPV was between 60% to 70% for all ESAs analyzed (Table 1). The sensitivity and specificity were considerably more variable (Table 1), as would be expected, identifying that two ESAs had a high sensitivity (30, 34), and another a high specificity (21). Finally, the AUC analysis revealed values from 0.512 to 0.629 (Table 1).

The IR for each category of four of the analyzed ESAs did not statistically significantly vary ($P>.05$) (Fig. 1). However, the IR for the three categories of the aneuploidy risk classification ESA (30) statistically significantly varied ($P<.0001$), as

FIGURE 1



Implantation rates (IRs) of the embryo classification categories in each of the analyzed embryo selection algorithms (ESAs). Azzarello et al. (21): IR of embryos where pronuclear fading (PNf) occurred after 20.75 hpi (n = 832, 37.74%) and embryos that faded before 20.75 hpi (n = 145, 43.45%) ($P>.05$, Fisher's exact test). Basile et al. (37): IR of embryos classified as A (t3 = 34–40 hpi and cc2 9–12 hours, n = 453), B (t3 = 34–40 hpi, cc2 >9 or <12 hours, n = 231), C (t3 <34 or >40 hpi and cc2 = 9–12h, n = 173) and D (t3 <34 or >40 hpi and cc2 <9 or >12 hours, n = 120) with respective IR of 43.05%, 38.53%, 35.84%,and 25.83% ($P>.006$, Kruskal-Wallis test). Campbell et al. (38): IR for embryos classified as low risk (tSB <92.2 hpi and tB <122.9 hpi, n = 621), medium risk (tSB ≥96.2 and tB ≤122.9 hpi, n = 353) and high risk (tB ≥122.9 hpi, n = 3) of aneuploidy with respective IR of 43.80%, 28.61% and 33.33% ($P<.05$, Kruskal-Wallis test). Chamayou et al. (34): IR of those embryos where cc3 (t5–t3) occurred between 9.7–21 hours (n = 959, 11.11%) and those that did not (n = 18, 38.79%) ($P>.05$, Fisher's exact test). Cruz et al. (27): IR of embryos classified as A (t5 = 48.8–56.6 hpi and s2 ≤0.76 hours, n = 364), B (t5 = 48.8–56.6 hpi and s2 >0.76 hours, n = 140), C (t5 ≤48.8 *or* >56.6 hpi and s2 ≤0.76 hours, n = 353) and D (t5 <48.8 *or* >56.6 and s2 >0.76 hours, n = 120) with respective IR of 41.21%, 30.71%, 39.94% and 35.83% ($P>.05$, Kruskal-Wallis test). Dal Canto et al. (35): IR of embryos where t8 occurred between 51.6–70.4 hpi (n = 578, 40.66%) and those that did not (n = 399, 35.59%) ($P>.05$, Fisher's exact test). cc2 = time to complete the second cell cycle; cc3 = time to complete the third cell cycle; hpi = hours postinsemination; s2 = time between three and four cell; t2 = time to two cell; t3 = time to three cell; t4 = time to four cell; t5 = time to five cell; t8 = time to eight cell; tB = time to blastocyst (when the blastocele has filled over half of the embryo and there is a <10% increase in the embryo diameter); tPNf = time to pronuclear fading; tSB = time to start of blastoculation.

Barrie. Efficacy of six published ESAs. Fertil Steril 2016.

did category A with category D in the ESA developed by Basile et al. (37). The aneuploidy risk classification ESA also had the strongest LR (1.26) and PPV (44.28%). Incidentally, the number of embryos classified as high risk using this ESA was just three, of which one implanted, giving this category an IR of 33.33%—a potentially misleading result. The absolute difference between the IR of low- and medium-risk embryos was 15.19% (Fig. 1).

DISCUSSION

All six of the examined ESAs (21, 27, 30, 34, 35, 37) achieved an AUC <0.65 (0.584, 0.558, 0.573, 0.612, 0.543, and 0.629, respectively), indicating reduced predictive capability. None of the ESAs achieved a PPV above 45% (42.57%, 41.52%, 44.28%, 38.91%, 38.29%, and 40.45%, respectively), which was also indicative of poor diagnostic value. The NPV reached over 60% in all the ESAs (62.12%, 68.26%, 71.34%, 76.19%, 61.10%, and 64.14%, respectively). The specificity of each ESA was variable, indicating that some ESAs are able to identify embryos with a reduced chance of implantation better than others (85.83%, 33.33%, 42.33%, 2.67%, 48.17%, and 42.33%, respectively) also reflected in the NPV. This variability was also, inevitably, seen in the sensitivity of the assessed ESAs (16.71%, 75.33%, 72.94%, 98.67%, 51.59%, and 62.33%, respectively).

In all the ESAs assessed, the LR was close to 1 (1.18, 1.13, 1.26, 1.01, 0.99, and 1.08, respectively). The LRs of all ESAs revealed that there was little predictive power of implantation where a favorable ESA result is obtained (Table 1). Likelihood ratios range from 0 to infinity and a LR close or equal to 1 indicates a lack of diagnostic value; the farthest from 1 that any of the ESAs in this investigation reached was 0.26, indicating that an embryo has a 0.26 increased chance of creating a pregnancy if a favorable ESA outcome is achieved.

Worthy of note are the ESAs that were found to have statistical significance between the categories of embryo classification (30, 37). However, the number of embryos classified as high risk of aneuploidy in the aneuploidy risk classification ESA was just three of 977. Further validation, performed by the developers of this ESA (38) using 88 embryos, classified four as high risk of aneuploidy. Clearly, using this ESA, the chance of an embryo being classified as high risk is low, which raises issues about the specificity of the ESA especially when evidence suggests that over 50% of embryos exhibit aneuploidy (39). With an AUC of 0.575 and a 0.26 increased chance that an embryo would create a pregnancy if classified as low risk of aneuploidy, this ESA may not represent a robust, clinically applicable embryo selection. Nonetheless, this ESA is the most effective out of the six assessed when a combination of specificity, sensitivity, PPV, NPV, AUC, LR, and differences in implantations between each embryo classification category is considered.

The other ESA to gain statistical significance between the categories when considering IR was that of Basile et al. (37). Statistical significance was found between the IRs of category A and D, indicating that this ESA may perform well in terms of identification of poor quality embryos. This is also reflected in a high sensitivity and NPV. However, the LR remains low at 1.13, and the other measures of the effectiveness of the ESA (specificity, PPV and AUC) indicate this ESA may not be as effective at determining higher implantation potential embryos.

The analyses performed indicate that ESAs available in the literature may not provide substantial, additional aid for embryo selection in a clinically relevant setting. The current investigation highlights that externally derived ESAs are developed, unavoidably, under conditions different to that of the adoptive center (Table 2) encouraging the development of in-house, specific ESAs. It has been shown that the method by which embryos are created (IVF or ICSI) can affect their temporal behavior (41–43). In addition to varying treatment types, a number of the analyzed ESAs excluded certain patient groups to avoid confounding factors. This includes those with endometriosis, polycystic ovary syndrome (PCOS), severe male factor infertility, and maternal age over 39 years. This exclusion constitutes a proportion of patients that make up a significant fraction of patients treated in an IVF laboratory and for which these ESAs could be critically useful.

There is evidence to suggest that the reason for infertility could affect an embryo's morphokinetic profile, in particular those with PCOS (44), thus their exclusion in the ESA development is understandable but reduces its clinical applicability unless a specific ESA is developed for this specific patient group. Furthermore, one group's ESA was developed using oocyte donors only, a clear confounder for the application of this ESA in other centers.

In addition, the majority of the ESAs were developed on embryos created under an agonist protocol. However, one group's ESA development cohort contained a proportion of embryos created under an antagonist protocol (30) The use of agonist and antagonist protocols has yet to be shown to affect an embryo's morphokinetic profile however they have been linked to embryo quality (45, 46), which could indicate that there is a potential for them to also have a temporal effect.

Perhaps most importantly, varying culture conditions were used in the development of these ESAs. It has been shown that an embryos morphokinetic profile is greatly altered in different culture media, specifically between sequential and single-step media (47, 48). This means that ESAs developed using sequential media may not be effective in selecting embryos cultured in single-step media, and vice versa. In addition, varying CO_2 and O_2 gas concentrations were used in the development of a number of these published ESAs. Oxygen tension has been specifically linked to an embryo's morphokinetic profile in both humans (49) and mice (40) where those embryos cultured at 20% O_2 have reduced developmental rates and the completion of the third cell cycle is significantly delayed. Of the six ESAs analyzed, one comprised multiple centers (37). The culture conditions varied slightly between centers, so it could be argued that this ESA has a broader clinical use while maintaining similar predictive power measurements (i.e., sensitivity, specificity, PPV, NPV, LR, AUC) to the other ESAs investigated. It should be highlighted however, that the algorithm developed in this original article used oocyte donors, a natural bias for outcomes focusing on embryologic features and implantation potential. These fundamental differences in the development of each ESA need to be seriously considered before their external adoption. It is highly unlikely that an external center

TABLE 2

Summary of publications used for examination of efficacy of selection criteria.

Study	Embryos, n	Cycles, n	Fertilization method	End point	Exclusion criteria	Inclusion criteria	Image capture interval(min)	Protocol	Culture	Media change	Transfer day
Azarello et al. 2012 (21)	159	130	ICSI	LBR	—	Embryos transferred at four-cell stage with equal blastomeres and <25% fragmentation, autologous gametes, female age ≤39, male factor infertility (1–5 × 10^5 motile sperm/ejaculate)	20	Agonist	Cook, 5.5% CO_2, 5% O_2, 89.5% N_2	No	2 (44 hpi)
Basile et al. 2015 (37)	754	1664	ICSI	IR	Severe male factor, severe endometriosis, BMI 30 kg/m^2, low response (for standard patients, less than five metaphase II oocytes) and no PGS or PGD	For donors: 18–30 y old (mean: 26.9, SD 4.7), normal menstrual cycles (26–34 d' duration), a BMI of 18–28 kg/m^2 and normal ovaries and uterus as observed by transvaginal ultrasound	10–20	Antagonist (agonist trigger for analogous oocytes, hCG trigger for oocyte donors)	Site 1: Cook, 5.5% CO_2, 20% O_2, 74.5% N_2 Site 2: Global IVF Medium, 6.5% CO_2, 20% O_2, 73.5% N_2 Site 3: Cook, 6% CO_2, 20% O_2, 74% N_2 Site 4: Global IVF Medium, 6% CO_2, 20% O_2, 74% N_2	No	3
Campbell et al. 2013 (30)	88	25	ICSI	CPR and LBR	—	Patients undergoing a cycle inclusive of PGS	20	Agonist (75%) Antagonist (25%)	Global IVF medium, 5.5% CO_2, 5% O_2, 89.5% N_2	Yes	
Chamayou et al. 2013 (34)	178	78	ICSI	BFR	Severe endometriosis, premature ovarian failure, severe asthenoteratozoospermia	Fresh gametes	20	Agonist	Quinn's Advantage, 5% CO_2, 5% O_2	Yes	5
Cruz et al. 2012 (27)	834	165	ICSI	BFR	—	Oocyte donor meeting all required criteria for donation program	20	Agonist	Global IVF medium, 6% CO_2, 21% O_2, 37.4°C	Yes	5

Barrie. Efficacy of six published ESAs. Fertil Steril 2016.

TABLE 2

Continued.

Study	Embryos, n	Cycles, n	Fertilization method	End point	Exclusion criteria	Inclusion criteria	Image capture interval(min)	Protocol	Culture	Media change	Transfer day
Dal Canto et al. 2012 (35)	134	71	IVF (22 cycles) and ICSI (49 cycles)	IR	—	Indication for standard IVF or ICSI due to male factor, tubal factor, stage I or II endometriosis, or PCOS, maternal age 27–42 y.	20	Agonist	ISM1 (day 1–3), BlastAssist (day 3–5), 6% CO_2, 5% O_2, 89% N_2	Yes	3 and 5

Note: BFR = blastocyst formation rate; CO_2 = carbon dioxide; CPR = clinical pregnancy rate; hpi = hours post insemination; ICSI = intracytoplasmic sperm injection; IR = implantation rate; IVF = in vitro fertilization; LBR = live birth rate; N_2 = nitrogen; O_2 = oxygen; PCOS = polycystic ovary syndrome; PGD = preimplantation genetic diagnosis; PGS = preimplantation genetic screening.

Barrie. Efficacy of six published ESAs. Fertil Steril 2016.

will have the same patient, treatment and environmental parameters as that of the developing center.

A further consideration for the use of externally derived ESAs is the subjective nature of annotating morphokinetic parameters, the differences in image-capture analysis such as the number of focal planes, and the varying definition of t0. The subjective nature of annotations creates unreliability in the external application of ESAs. There has been some development with this due to the publication of annotation guidelines in 2014 (50), but this will not eliminate the subjectivity entirely. It is interesting that there are now two commercially available one-size-fits-all ESAs that, based on the results presented here, should not perform as well as expected. Variations in image acquisition are unlikely to create significant disparity, but coupled with the variability between "annotators," an increasing level of inaccuracy could be created. Although it is undefined in some of the publications, the definition of t0 varies between groups; some use t0 as the time of insemination or injection, the inaugural and arguably the most common method, and others use it as the midpoint of ICSI. It has now been largely accepted that the use of insemination/injection is arbitrary, the exact moment that the sperm enters the oocyte is indeterminate for IVF cases, and where possible, the time of pronuclei fading should be used as t0.

It could be argued that a limitation of the current analyses is the potential for bias due to the use of an in-house ESA with similar morphokinetic parameters to one of the externally derived ESAs (27) to aid in embryo selection of the analyzed embryos. Owing to this, we compared the proportion of embryos in each of the categories (A–D) in the original manuscript for the external ESA in question (27) with the current analyses. From this analysis, the proportion of embryos in each category did not differ between the original manuscript of the external ESA and the current analyses: A, 39.7% (106 of 267) versus 37.3% (364 of 977); B, 13.5% (36 of 267) versus 14.3% (140 of 977); C, 36.0% (96/267) versus 36.1% (353 of 977); and D, 10.8% (29 of 267) versus 12.3% (120 of 977), respectively. This provides reassurance that any bias created from the use of similar morphokinetic parameters in the selection of the embryos used in this analysis is minimal.

Finally, it is important to consider that the use of time-lapse imaging as a method for embryo selection has yet to be appropriately evidenced (51). As can be seen from the results presented here, the poor performance of the investigated ESAs allows the field to question the overall clinical applicability of the use of time-lapse systems. There is considerable heterogeneity in the origin and culture of the embryos used for the development of these ESAs, and it should be considered that these parameters affect the ability of a one-size-fits all approach to function effectively. Perhaps the development of optimum morphokinetic time ranges that are patient, treatment, and environment specific will present a means of using time-lapse systems to achieve a higher predictive power.

There are ideal conditions under which to test the efficacy of externally derived ESAs; select embryos based only on morphology, then perform the analyses presented here or, preferably, prospectively apply ESAs. At the study site, morphokinetics have been used since their introduction in the laboratory to aid in embryo selection, so a data set large enough to perform

the former of these two methodologies would not be possible. We do, however, recognize the strength of a prospective methodology for the aims presented here. This will be the focus of future research in this area to better delineate the benefits of using time-lapse systems in the clinical embryology laboratory.

CONCLUSION

The development of ESAs thus far has not involved the control of confounding factors such as media type, patient age, and treatment type, except inadvertently by virtue of availability. They are often developed under the environmental parameters available in the laboratory performing the development and thus are clinically relevant in these cases alone. For external application, the ESAs lose their predictive capabilities.

The primary objective of ESAs is to allow the selection of the best embryo from a cohort in a clinical setting. Those presented here clarify that embryo morphokinetics could be used for embryo selection, but they do not offer a clinically relevant means to aid in embryo selection in other laboratories unless the development criteria are also adopted. The collective contribution of confounding factors means that derived ESAs can only be applied to the conditions under which they were developed; when applied to a heterogeneous cohort of embryos, as would be found in an IVF laboratory, the capability of the ESA to detect the most viable embryo diminishes. Further research needs to focus on the development of ESAs that are specific to subgroups of patients, environments, and treatments. At the very least, embryology laboratories should proceed with caution when implementing ESAs derived from published sources and consider thorough in-house validation of such ESAs before clinical use, if ESAs are to be used at all.

REFERENCES

1. Tesarik J, Greco E. The probability of abnormal preimplantation development can be predicted by a single static observation on pronuclear stage morphology. Hum Reprod 1999;14:1318–23.
2. Scott L. Pronuclear scoring as a predictor of embryo development. Reprod Biomed Online 2003;6:201–14.
3. Payne D, Flaherty SP, Barry MF, Matthews CD. Preliminary observations on polar body extrusion and pronuclear formation in human oocytes using time lapse video cinematography. Hum Reprod 1997;12:532–41.
4. De Placido G, Wilding M, Strina I, Alviggi E, Alviggi C, Mollo A, et al. High outcome predictability after IVF using a combined score for zygote and embryo morphology and growth rate. Hum Reprod 2002;17:2402–9.
5. Palmstierna M, Murkes D, Csemiczky G, Andersson O, Wramsby H. Zona pellucida thickness variation and occurrence of visible mononucleated blastomeres in pre embryos are associated with a high pregnancy rate in IVF treatment. J Assist Reprod Genet 1998;15:70–5.
6. Lundin K, Bergh C, Hardarson T. Early embryo cleavage is a strong indicator of embryo quality in human IVF. Hum Reprod 2001;16:2652–7.
7. Isiklar A, Mercan R, Balaban B, Alatas C, Aksoy S, Urman B. Early cleavage of human embryos to the two cell stage. A simple, effective indicator of implantation and pregnancy in intracytoplasmic sperm injection. J Reprod Med 2002;47:540–4.
8. Pickering SJ, Taylor A, Johnson MH, Braude PR. An analysis of multinucleated blastomere formation in human embryos. Hum Reprod 1995;10:1912–22.
9. Jackson KV, Ginsburg ES, Hornstein MD, Rein MS, Clarke RN. Multiniucleation in normally fertilized embryos is associated with an accelerated ovulation induction response and lower implantation and pregnancy rates in in vitro fertilization embryo transfer cycles. Fertil Steril 1998;70:60–6.
10. Yakin K, Balaban B, Urman B. Impact of the presence of one or more multinucleated blastomeres on the developmental potential of the embryo to the blastocyst stage. Fertil Steril 2005;83:243–5.
11. Shapiro BS, Harris DC, Richter KS. Predictive value of 72 hour blastomere cell number on blastocyst development and success of subsequent transfer based on the degree of blastocyst development. Fertil Steril 2000;73:582–6.
12. Hardarson T, Hanson C, Sjogren A, Lundin K. Human embryos with unevenly sized blastomeres have lower pregnancy and implantation rates: indications for aneuploidy and multinucleation. Hum Reprod 2001;16:313–8.
13. Johansson M, Hardarson T, Lundin K. There is a cutoff limit in diameter between a blastomere and a small anucleate fragment. J Assist Reprod Genet 2003;20:309–13.
14. Freour T, Lammers J, Splingart C, Jean M, Barriere P. Time lapse (EmbryoScope as a routine technique in the IVF laboratory: a useful tool for better embryo selection? Gynecol Obstet Fertil 2010;40:476–80.
15. Nakahara T, Iwase A, Goto M, Harata T, Suzuki M, Ienaga M, et al. Evaluation of the safety of time lapse observations for human embryos. J Assist Reprod Genet 2010;27:93–6.
16. Basile N, Morbeck D, Garcis-Velasco J, Bronet F, Meseguer M. Time lapse technology reveals that embryo kinetics are not affected by culture media [abstract]. Fertil Steril 2011;96(Suppl):S108.
17. Cruz M, Gadea B, Garrido N, Pedersen KS, Martinez M, Perez-Cano I, et al. Embryo quality, blastocyst and ongoing pregnancy rates in oocyte donation patients whose embryos were monitored by time lapse imaging. J Assist Reprod Genet 2011;28:569–73.
18. Kirkegaard K, Hindkjaer JJ, Grondahl ML, Kesmodel US, Ingerslev HJ. A randomized clinical trial comparing embryo culture in a conventional incubator with a time lapse incubator. J Assist Reprod Genet 2012;29:565–72.
19. Lemmen JG, Agerholm I, Ziebe S. Kinetic markers of human embryo quality using time lapse recordings of IVF/ICSI fertilized oocytes. Reprod Biomed Online 2008;17:385–91.
20. Scott L. The origin and consequences of day 2 multinucleation of human embryos [abstract]. Abstracts of the 26th annual meeting of ESHRE 2010, Rome, Italy. Hum Reprod 2010;25(Suppl 1):204.
21. Azzarello A, Hoest T, Mikkelsen AL. The impact of pronuclei morphology and dynamicity on live birth outcome after time lapse culture. Hum Reprod 2012;27:2649–57.
22. Gonzales DS, Pinheiro JC, Bavister BD. Prediction of the developmental potential of hamster embryos in vitro by precise timing of the third cell cycle. J Reprod Fertil 1995;105:1–8.
23. Ramsing NB, Cellesen H. Detecting timing and duration of cell divisions by automatic image analysis may improve selection of viable embryos [abstract]. Fertil Steril 2006;86(Suppl):S189.
24. Ramsing NB, Berntsen J, Callesen H. Automated detection of cell division and movement in time lapse images of developing bovine embryos can improve selection of viable embryos [abstract]. Fertil Steril 2007;88(Suppl 1):S38.
25. Lechniak D, Pers-Kamczyc E, Pawlak P. Timing of the first zygotic cleavage as a marker of developmental potential of mammalian embryos. Reprod Biol 2008;8:23–42.
26. Herrero J, Tejera A, Ramsing N, Romero JL, Rubio I. Establishing the optimal time ranges of key events during development using time lapse video cinematography [abstract]. Fertil Steril 2011;96(Suppl):S102.
27. Cruz M, Garrido N, Herrero J, Perez-Cano I, Munoz M, Meseguer M. Timing of cell divisions in human cleavage stage embryos correlates with blastocyst formation and quality. Reprod Biomed Online 2012;25:371–81.
28. Hlinka D, Kalatova B, Uhrinova I, Dolinska S, Rutarova J, Rezacova J. Time lapse cleavage rating predicts human embryo viability. Physiol Res 2012;61:513–25.
29. Meseguer M, Rubio I, Cruz M, Basile N, Marcos J, Requena A. Embryo incubation and selection in a time lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. Fertil Steril 2012;98:1481–9.
30. Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Hickman CF. Modelling a risk classification of aneuploidy in human embryos using non invasive morphokinetics. Reprod Biomed Online 2013;26:477–85.

31. Hardarson T, Lofman C, Coull G, Sjogren A, Hamberger L, Edwards RG. Internalization of cellular fragments in a human embryo: time lapse recordings. Reprod Biomed Online 2002;5:36–8.
32. 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.
33. Liu Y, Chapple V, Roberts P, Matson P. Prevalence, consequence, and significant of reverse cleavage by human embryos viewed with the use of the EmbryoScope time lapse video system. Fertil Steril 2014;102:1295–300.
34. Chamayou S, Patrizio P, Storaci G, Tomaselli V, Alecci C, Ragolia C, et al. The use of morphokinetic parameters to select all embryos with full capacity to implant. J Assist Reprod Genet 2013;30:703–10.
35. Dal Canto M, Coticchio G, Mignini Renzini M, De Ponti E, Novara PV, Brambillasca F, et al. Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation. Reprod Biomed Online 2012; 25:474–80.
36. Cutting R, Morroll D, Roberts SA, Pickering S, Rutherford A, BFS and ACE. Elective single embryo transfer: guidelines for practice British Fertility Society and Association of Clinical Embryologists. Hum Fertil 2008;11:131–46.
37. 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.
38. Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time lapse imaging without PGS. Reprod Biomed Online 2013;27:140–6.
39. Fragouli E, Wells D. Aneuploidy in the human blastocyst. Cytogenet Genome Res 2011;133:149–59.
40. Wale PL, Gardner DK. Time lapse analysis of mouse embryo development in oxygen gradients. Reprod Biomed Online 2010;21:402–10.
41. Cruz M, Garrido N, Gadea B, Munoz M, Perez-Cano I, Meseguer M. Oocyte insemination techniques are related to alterations of embryo developmental timing in an oocyte donation model. Reprod Biomed Online 2013;27:367–75.
42. Bodri D, Sugimoto T, Serna JY, Kondo M, Kato R, Kawachiya S, et al. Influence of different oocyte insemination techniques on early and late morphokinetic parameters: retrospective analysis of 500 time lapse monitored blastocysts. Fertil Steril 2015;104:1175–81.
43. Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Time lapse videography of human embryos: using pronuclear fading rather than insemination in IVF and ICSI cycles removes inconsistencies in time to reach early cleavage milestones. Reprod Biol 2015;15:122–5.
44. Wissing ML, Bjerge MR, Olesen AI, Hoest T, Mikkelsen AL. Impact of PCOS on early embryo cleavage kinetics. Reprod Biomed Online 2014;28:508–14.
45. Murber A, Fancsovits P, Ledo N, Gilan ZT, Rigo J, Urbancsek J. Impact of GnRH analogues on oocyte/embryo quality and embryo development in in vitro fertilization/intracytoplasmic sperm injection cycles: a case control study. Reprod Biol Endocrinol 2009;25:103.
46. Vengetesh PM, Ramachandran A, Kumar P. Choosing GnRH Antagonist protocol shows improved oocyte and embryo quality, coherent with perifollicular vascularity (PFV) in assisted reproductive techniques. J Clin Diagn Res 2015;9:QC24–8.
47. 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.
48. Barrie A, Taylor E, Schnauffer K, Kingsland C, Troup S. An examination of embryo morphokinetics and utilisation in single step and sequential culture media systems [abstract]. Hum Reprod 2015;30(Suppl 1):282.
49. 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.
50. 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.
51. Kaser DJ, Racowsky C. Clinical outcomes following selection of human preimplantation embryos with time lapse monitoring: a systematic review. Hum Reprod Update 2014;20:617–31.