**human
reproduction
open**

ORIGINAL ARTICLE

High reliability of morphokinetic annotations among embryologists

M. Martínez[1], J. Santaló[2], A. Rodríguez[1], and R. Vassena ⓘ [1,*]

[1]Clínica Eugin, Travessera de les Corts 322, Barcelona 08029, Spain [2]Departamento de Biología Celular, Fisiología en Inmunología, Facultad de Biociencias, Universidad Autónoma de Barcelona, Bellaterra 08193, Cerdanyola del Vallès, Barcelona

*Correspondence address. Clínica Eugin, Travessera de les Corts 322, Barcelona 08029, Spain. Tel: +34-93-322-1122; E-mail: rvassena@ eugin.es ⓘ orcid.org/0000-0002-0846-0365

STUDY QUESTION: Are morphokinetic measurements of time lapse-videos of human embryos comparable among operators?

SUMMARY ANSWER: There is little variation among morphokinetic measurements taken by different operators when analyzing the same time lapse-videos of human embryos.

WHAT IS KNOWN ALREADY: Morphokinetic analysis of preimplantation embryo development is a complementary method of embryo assessment increasingly used in IVF laboratories. Time-lapse videos of embryo development are normally viewed by trained embryologists and annotated with the times when specific developmental events occur. Such annotations form the basis of embryo selection algorithms, used to rank the embryos for transfer. It is unknown whether the reliability of morphokinetic annotations is related to the morphological characteristics of the analyzed embryo or to the ability of the embryologists performing the annotation. One study so far reported the reliability of morphokinetic annotations among three embryologists using the time-lapse system (TLS), but larger studies with different setups are needed to address this issue further.

STUDY DESIGN, SIZE, DURATION: A prospective study was carried out between October 2015 and June 2016. Six embryologists with various degrees of experience in static, morphology-based evaluation, individually annotated the same 93 videos of preimplantation development, corresponding to 18 IVF/ICSI cycles, recorded with a TLS.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Times of second polar body extrusion, appearance and disappearance of pronuclei, and embryo cleavages (times from 2-cell to 5-cell stage: t2, t3, t4, t5) were annotated. Each embryologist was blinded to the annotations of the others. Intra- and inter-observer agreement was evaluated by computing intra-class correlation coefficients (ICCs).

MAIN RESULTS AND THE ROLE OF CHANCE: In the inter-observer analysis, most ICCs obtained were higher than 0.80, indicating a high level of agreement: t2: 0.93; t3: 0.80; t4: 0.89; t5: 0.89; disappearance of two pronuclei: 0.98. However, the ICCs obtained for second polar body extrusion and the appearance of two pronuclei annotations was lower: 0.51 and 0.63, respectively, indicating an average level of agreement. The ICCs obtained from the intra-observer analysis were also higher than 0.80 (t2: 0.96; t3: 0.89; t4: 0.88; t5: 0.86; disappearance of two pronuclei: 0.96). The ICCs obtained from second polar body extrusion and the appearance of two pronuclei annotations were 0.77 and 0.66, respectively. These results indicate that developmental timings, annotated in time-lapse videos, are highly reliable both within and among observers.

LIMITATIONS, REASONS FOR CAUTION: The events at the developmental stages from 6-cells to blastocyst were not evaluated; since some morphokinetic algorithms use times past the 6-cell stage in their calculations, further studies should be carried out to understand the variations among observers in these cases.

WIDER IMPLICATIONS OF THE FINDINGS: Time-lapse measurement should be as objective as possible, especially for the first embryo cleavages, because they are often measured to define algorithms to assess the embryonic implantation potential. Our results show that measurements using this particular TLS are consistent and reliable both within and among operators.

STUDY FUNDING/COMPETING INTEREST(S): None.

TRIAL REGISTRATION NUMBER: Not applicable.

WHAT DOES THIS MEAN FOR PATIENTS?

Time-lapse imaging is now used in many IVF laboratories around the world to see how embryos develop and to select the best embryo to transfer. The images are assessed by individual embryologists and this paper looks at whether there is variation in the way the embryologists make their assessments.

The researchers took 93 videos of embryo development from 18 IVF/ICSI cycles and got six embryologists, four senior and two junior, to mark a number of stages of development. They then compared the points at which they had identified the changes in the embryo to see whether there was variation between the different embryologists. They also looked at whether a single embryologist would make the same analysis when looking at the video again.

The results showed that the embryologists were consistent in the way they analyzed embryos using time lapse but there were anomalies in identifying one early stage of development. The researchers suggest that further studies would be useful to investigate possible variability.

Introduction

Time-lapse systems (TLS) have been introduced in many IVF laboratories around the world in the last few years, in order to analyze embryo morphokinetic development with the objective to increase the pregnancy rates in IVF cycles (Meseguer et al., 2011; Storr et al., 2015; Motato et al., 2016) by selecting the embryos with the highest implantation potential for transfer. In addition to providing further information about embryonic development, TLS allows for culturing the embryos undisturbed in the incubator. Nevertheless, recently TLS diagnostic usefulness has been questioned by several authors (Polanski et al., 2014; Armstrong et al., 2015a, 2015b; Kirkegaard et al., 2015; Racowsky et al., 2015). Although the time-lapse analysis seems to be a simple task, it remains so far a subjective assessment, which could lead to variability in the morphokinetic evaluation of embryos. To standardize this field, proposals for a standard nomenclature to be used in the TLS (Ciray et al., 2014) as well as intra- and inter-observer variability checks of the morphokinetic parameters of embryonic development annotations using Embryoscope® (Sundvall et al., 2013) have been put forward. These are important because both intra- and inter-observer heterogeneity could bias the annotations, potentially impacting the morphokinetic analysis results. Most TSL annotations are still embryologist-dependent and manually recorded, and therefore can be biased by the subjectivity of the embryologist. The aim of this study is to evaluate the intra- and inter-observer reliability of the morphokinetic annotations of embryonic development, and assess whether embryology experience and/or embryo morphology play a role in its definition.

Materials and Methods

Study design and ethical approval

This is a prospective study performed between October 2015 and June 2016, where six embryologists evaluated the videos of 93 developing human embryos from 18 IVF cycles. The cycles included were randomly selected among all cycles performed within the study period. Written informed consent was obtained from each embryologist participating in the study. Ethical approval from the Ethics Committee for Clinical Research of Clínica EUGIN was obtained before the implementation of the study.

Characteristics of embryologists

Six clinical embryologists participated in the study; four senior embryologists (A–D) each with 6–11 years of experience in conventional embryo morphology-based evaluation, and the two junior embryologists in the laboratory team (E and F) each with 1–2 years of experience. The senior embryologists (A–D) had performed more than 2000 embryo transfer (ET) selections and morphological evaluations throughout their career, while juniors (E and F) had performed around 400 ET selections each. Except for embryologist A, none had ever performed an embryo evaluation with TLS before their participation in the study. The demographic characteristics of the embryologists are described in Table I.

Characteristics of the analyzed embryos

The 93 embryos included in the study were derived from oocyte donation cycles fertilized by ICSI with either partner or donor sperm. Oocyte donors were 23–29 years old and all semen samples used were

Table I Demographic characteristics of the participant embyologists.

	Age (years)	Working experience (years)
Senior embryologists		
A	34	10
B	32	6
C	38	13
D	32	10
Junior embryologists		
E	26	2
F	25	2

normozoospermic ([World Health Organization 2010](#)). Embryo morphological score and cycle reproductive outcomes were not taken into account for inclusion, so the embryo analysis was carried out on embryos with differing morphology and implantation potential. Morphological embryo quality of the embryos ranged from 1 to 10 ([Coroleu et al., 2006](#)).

Primo Vision® System procedures

Primo Vision® (Vitrolife, Göteborg, Sweden) captured a bright field image of the embryos every 5 min, and an additional scan (11 focal planes over a 100-μm scan range) every 20 min. The two Primo Vision® microscopes used in the study were placed in incubator at 37°C, 6% CO_2, and 95% relative humidity. Embryo development videos were recorded by Primo Vision Capture®, and the time-lapse operators annotated the videos using Primo Vision Analyzer® (Vitrolife, Göteborg, Sweden).

An initial theoretical class was provided to all participating embryologists with the aim of standardizing the knowledge of TLS among the operators. This class was provided by the main investigator (embryologist A) and lasted 90 min. Topics discussed were the operation of the Primo Vision® System, how to perform embryo culture using TLS, and how to record and annotate developmental events.

Morphokinetic data collection

The morphokinetic timing nomenclature was revised before the start of data collection, based on the guidelines of [Ciray et al. (2014)](#). Time from ICSI (t0) to each developmental event was calculated as the time when the first oocyte of the cohort was injected (the injection of the entire cohort was completed in <10 min in all cases), and each developmental event was determined for each time-lapse recording. The analyzed events were extrusion of the second polar body (tPB2), pronuclei appearance (tPN), pronuclear breakdown (tPBf), and division to 2-cell through 5-cell stages (t2–t5). These events are described in Table II.

Statistical analysis

Inter-observer reliability of TLS measurements was performed using the assessments of each embryologist A–F. The intra-observer reliability was performed using two assessments of the embryologist A: the initial one and an additional measurement repeated 6 months later.

Inter-observer and intra-observer intra-class correlation coefficients (ICC) for single measures using an absolute agreement were calculated for each proposed morphokinetic parameter. The ICC values were interpreted according to [Shrout and Fleiss (1979)](#) as: >0.8 as a high–perfect agreement; 0.7–0.8 as a good agreement; 0.5–0.6 as a regular–moderate agreement; 0.2–0.4 as poor–fair agreement.

In addition, the inter-observer analysis was stratified in two groups according to the working experience of the embryologist (senior or junior) and in three groups of embryo morphological quality according to [Coroleu et al. (2006)](#). High quality was defined as embryos with a score from 8 to 10, medium quality as embryos with score of 7, and low quality as embryos with a quality equal to or lower than 6.

Results

Inter-observer analysis

In the inter-observer analysis, most ICCs obtained were higher than 0.80, indicating a high level of agreement among embryologists, with the highest value being 0.98 for pronuclei breakdown (tBPN). However, the ICCs obtained for the second polar body extrusion (t2PB) and the appearance of two pronuclei (tPN) were lower, at 0.51 and 0.63, respectively, indicating a moderate agreement. The 3-cell stage was recorded by all embryologists in just 14 embryos. We think that this could be due to the short duration of the 3-cell stage, and to the fact that embryologists were unwilling to annotate an event if they felt unsure about its exact timing. The complete results for the inter-observer analysis are described in Table III.

The stratified analyses showed that the ICCs obtained do not change with either the experience in morphology-based embryo scoring of the embryologist (Table IV), or with the morphological score of the annotated embryos (Table V).

Intra-observer analysis

Most ICCs obtained from the intra-observer analysis were higher than 0.80. The lowest value was obtained from t2PB and tPN, at 0.77 and 0.66, respectively, indicating a good agreement. The results for the intra-observer analysis are presented in Table VI.

Discussion

Embryo quality is a determining factor in IVF success and is usually evaluated through observation of embryo morphology at fixed times ([Coroleu et al., 2006](#)). TLS allows the continuous monitoring of development and a determination of the exact times of embryological events ([Meseguer et al., 2011, 2012](#); [Petersen et al., 2016](#); [Conaghan et al., 2013](#); [Paternot et al., 2011](#)).

Table II Description and definition of the morphokinetic parameter analyzed in human embryos.

Morphokinetic parameter	Description	Definition
tPB2	Extrusion of the second polar body	*Time when the second polar body is completely detached from the oolemma
tPN	Appearance of pronuclei	*Time when fertilization status is confirmed
tPBf	Pronuclear breakdown	*Time of pronuclei disappearance
t2	First cleavage division	Cleavage from zygote to 2-cell stage embryo
t3	Second cleavage division	Cleavage from 2-cell to 3-cell stage embryo
t4	Third cleavage division	Cleavage from 3-cell to 4-cell stage embryo
t5	Fourth cleavage division	Cleavage from 4-cell to 5-cell stage embryo

*Definitions based on the guidelines proposed by [Ciray et al. (2014).](#)

Table III Inter-observer intra-class correlation coefficient for single measures using an absolute agreement definition for the analyzed morphokinetic parameter in human embryos.

Morphokinetic parameter	ICC (95% CI)	Analyzed embryos (n)	Level of concordance
Extrusion of the second polar body (tPB2)	0.516 (0.381–0.661)	54	Regular–moderate
Appearance of pronuclei (tPN)	0.631 (0.541–0.720)	75	Moderate–good
Pronuclear breakdown (tBPN)	0.980 (0.972–0.986)	73	High–perfect
First cleavage division (t2)	0.933 (0.908–0.953)	74	High–perfect
Second cleavage division (t3)	0.804 (0.650–0.920)	14	High–perfect
Third cleavage division (t4)	0.892 (0.850–0.926)	63	High–perfect
Fourth cleavage division (t5)	0.891 (0.849–0.926)	52	High–perfect

ICC and level of concordance according the intra-class correlation coefficient (Shrout and Fleiss, 1979).

The inter-observer intra-class correlation coefficient (ICC) values were interpreted according Shrout and Fleiss (1979), as: >0.8 as a high–perfect agreement; 0.7–0.8 as a good agreement; 0.5–0.6 as a regular–moderate agreement; 0.2–0.4 poor–fair agreement.

Table IV Inter-observer ICC for single measures using an absolute agreement definition (ICCa) for analyzed morphokinetic parameters stratified according to embryologist experience in embryo morphological analysis.

Morphokinetic parameter	Senior embryologist analyzer		Junior embryologist analyzer	
	ICC (95% CI)	Analyzed embryos (n)	ICC (95% CI)	Analyzed embryos (n)
Extrusion of the second polar body (tPB2)	0.40 (0.43–0.57)	41	0.27 (0.06–0.46)	81
Appearance of pronuclei (tPN)	0.46 (0.81–0.62)	85	0.61 (0.45–0.79)	83
Pronuclear breakdown (tBPN)	0.90 (0.92–0.96)	83	0.98 (0.97–0.99)	83
First cleavage division (t2)	0.93 (0.91–0.96)	80	0.89 (0.83–0.093)	81
Second cleavage division (t3)	0.76 (0.56–0.094)	14	0.91 (0.87–0.94)	74
Third cleavage division (t4)	0.88 (0.83–0.92)	71	0.95 (0.88–0.95)	75
Fourth cleavage division (t5)	0.87 (0.82–0.92)	58	0.72 (0.57–0.82)	64

The ICC values were interpreted according to Shrout and Fleiss (1979) as: >0.8 as a high–perfect agreement; 0.7–0.8 as a good agreement; 0.5–0.6 as a regular–moderate agreement; 0.2–0.4 poor–fair agreement.

Table V Inter-observer ICCs for single measures using an absolute agreement definition (ICCa) for analyzed morphokinetic parameters stratified according to embryo morphological scores.

Morphokinetic parameters	High score embryos		Medium score embryos		Low score embryos	
	ICCa (95% CI)	n	ICCa (95% CI)	n	ICCa (95% CI)	n
Extrusion of the second polar body	0.67 (0.48–0.85)	14	0.017	12	0.74 (0.99-0.99)	12
Appearance of pronuclei	0.43 (0.24–0.648)	23	0.51 (0.33–0.69)	30	0.7 (0.53–0.97)	18
Pronuclear breakdown	0.96 (0.93–0.98)	23	0.98 (0.95–0.99)	31	0.96 (0.61–0.92)	15
First cleavage division	0.99 (0.99–1.0)	24	0.85 (0.77–0.91)	30	0.99 (0.54–0.96)	17
Second cleavage division	0.7 (0.17–1.0)	2	0.78 (0.515–0.95)	6	0.8 (0.54–0.90)	5
Third cleavage division	0.87 (0.78–0.93)	20	0.99 (0.99–0.99)	29	0.78 (0.53–0.85)	12
Fourth cleavage division	0.91 (0.85–0.96)	19	0.91 (0.86–0.95)	27	0.8 (0.92–0.98)	6

The embryo morphology scores were calculated according to Coroleu *et al.* (2006).

Embryo scores were stratified as: high score (8–10), medium score (7) and low score (lower than 7).

ICCa values were interpreted according Shrout and Fleiss (1979) as: >0.8 as a high to perfect agreement; 0.7–0.8 as a good agreement; 0.5–0.6 as a regular to moderate agreement; 0.2–0.4 poor to fair agreement.

Table VI Intra-observer ICC for single measures using an absolute agreement definition (ICCa) for the morphokinetic parameters analyzed.

Morphokinetic parameter	Inter ICCa (95% CI)	n	Level of concordance
Extrusion of the second polar body	0.774 (0.625–0.860)	84	Moderate–good
Appearance of pronuclei	0.667 (0.552–0.772)	89	Moderate–good
Pronuclear breakdown	0.965 (0.948–0.977)	89	High–perfect
First cleavage division	0.962 (0.931–0.977)	88	High–perfect
Second cleavage division	0.894 (0.840–0.930)	83	High–perfect
Third cleavage division	0.884 (0.825–0.924)	80	High–perfect
Fourth cleavage division	0.860 (0.771–0.915)	65	High–perfect

The ICCa values were interpreted according to Shrout and Fleiss (1979) as: >0.8: high–perfect agreement; 0.7–0.8: good agreement; 0.5–0.6: regular–moderate agreement; 0.2–0.4: poor–fair agreement.

However, embryologist subjectivity is known to inform the morphology-based assessments when performed at fixed times, thereby affecting the selection for ET (Arce *et al.*, 2006; Baxter Bendus *et al.*, 2006; Paternot *et al.*, 2011). In the same way, subjectivity could affect TLS evaluations, as morphokinetic analysis is largely based on operator annotations with the exception of those TLS that provide automated analysis (Wong *et al.*, 2010, 2013).

Since most embryo selection algorithms are based on the timings of the early cell cycles (Meseguer *et al.*, 2011, 2012; Petersen *et al.*, 2016; Conaghan *et al.*, 2013; Paternot *et al.*, 2011), we took into account the annotations from zygote through the 5-cell stage.

Our results show a good agreement for the first cleavage annotations within embryologists, reaching an ICC of 0.80, similar to that reported by Shrout and Fleiss (1979), indicating that these annotations are reliable across different TLS. Moreover, our findings are in overall agreement with a previous study, reporting a good inter-observer concordance for TLS annotations (Sundvall *et al.*, 2013). In the same way, we stratified our analysis by the level of experience of the embryologists to test if the seniority (and hence experience) of the embryologist affects their morphokinetic evaluations. We show that the morphokinetic parameters associated with the timing of the embryo cleavages to the 5-cell stage, timings often used in TLS algorithms (Meseguer *et al.*, 2011), do not seem to be affected by the training of the embryologist. Furthermore, we assessed embryos of varying morphological scores, and we show that the inter-observer concordance does not change with the embryo score. Finally, we included a higher number of observers in comparison to the available literature, which results in a more robust analysis of inter-observer correlations.Independently from the possible and still debated use of these annotations to 'select in' embryos for transfer, TLS have an immediate use in 'selecting out' embryos with developmental alterations that might escape static observation. Events such as the tPB2, tPN, and tPNB are particularly critical in order to identify, for instance, abnormal fertilization, and are therefore of independent relevance to the embryologist. We report a high inter- and intra-observer ICCs (>0.90) for tPNB variability analysis, indicating high reliability. In the tPN estimates variability analysis, the ICCs obtained decreased to 0.63 and 0.66 for inter and intra-observer reliability analysis, respectively, indicating a moderate–good degree of agreement within and between observers. This decrease may be due to the difficulty in interpreting the presence of the two pronuclei with the software used, combined with the inability to move the embryo around in order to obtain the clearest possible plane of vision.

For the variability analysis of tPN and tPNB, our results are in agreement with a previous report which showed a good level of concordance both within and among observers (Sundvall *et al.*, 2013). Nevertheless, no study so far has evaluated the variability of tPB2, and the presence of two polar bodies is another parameter to consider in the evaluation of correct fertilization. In the intra-observer analysis, the ICC obtained from this event was 0.77, indicating a good degree of reliability. But in the inter-observer analysis the ICC was 0.51, indicating a regular–moderate degree of concordance, which further decreased to 0.27 when the embryologists were juniors. These results indicate that the annotation of this event is affected by both embryologist subjectivity and the level of training. In general, the visualization of tPB2 extrusion in TLS is difficult compared to other events, such as the first cleavage, possibly related to the second polar body size, the slightly inferior quality of the image, which is not as crisp as in some inverted microscopes, and the difficulty in placing the focus on the appropriate plane to observe the polar body.

We think that these results are relevant because, although not described in morphokinetic algorithms, tPB2 has been associated with embryonic aneuploidy (Chawla *et al.*, 2015) and live birth outcomes (Azzarello *et al.*, 2012).

Finally, although the ICCs obtained (except for tPB2) show a good reliability in the annotation of proposed events, our result with tPB2 indicate that we should be cautious in applying time-lapse imaging to analyze certain developmental events. We should be aware that some variability among embryologists exists and take this into account if we were to include such mophokinetic parameters in clinical decisions or selection algorithms.

We recognize some limitation in this study: the inclusion of two junior embryologists might have provided too small a sample, and we only evaluated preimplantation embryo development up to the third cell cycle. Time to compaction or time to blastocyst formation have not been calculated, hence, further studies will be needed to investigate the possible variability of these time-lapse annotations, as well as other morphokinetic parameters.

Conflict of interest

The authors have nothing to declare.

References

Arce JC, Ziebe S, Lundin K, Janssens R, Helmgaard L, Sorensen P. Interobserver agreement and intraobserver reproducibility of embryo quality assessments. *Hum Reprod* 2006;**21**:2141–2148.

Armstrong S, Arroll N, Cree LM, Jordan V, Farquhar C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev* 2015a;**2**:CD011320.

Armstrong S, Vail A, Mastenbroek S, Jordan V, Farquhar C. Time-lapse in the IVF-lab: how should we assess potential benefit? *Hum Reprod* 2015b;**30**:3–8.

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**:2649–2657.

Baxter Bendus AE, Mayer JF, Shipley SK, Catherino WH. Interobserver and intraobserver variation in day 3 embryo grading. *Fertil Steril* 2006;**86**: 1608–1615.

Chawla M, Fakih M, Shunnar A, Bayram A, Hellani A, Perumal V, Divakaran J, Budak E. Morphokinetic analysis of cleavage stage embryos and its relationship to aneuploidy in a retrospective time-lapse imaging study. *J Assist Reprod Genet* 2015;**32**:69–75.

Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, Sayed S, Time-Lapse User G. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod* 2014;**29**:2650–2660.

Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, Baker VL, Adamson GD, Abusief ME, Gvakharia M *et al.* Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril* 2013;**100**:412–419.e415.

Coroleu B, Barri PN, Carreras O, Belil I, Buxaderas R, Veiga A, Balasch J. Effect of using an echogenic catheter for ultrasound-guided embryo transfer in an IVF programme: a prospective, randomized, controlled study. *Hum Reprod* 2006;**21**:1809–1815.

Kirkegaard K, Ahlstrom A, Ingerslev HJ, Hardarson T. Choosing the best embryo by time lapse versus standard morphology. *Fertil Steril* 2015; **103**:323–332.

Meseguer M, Herrero J, Tejera A, Hilligsoe KM, Ramsing NB, Remohi J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod* 2011;**26**:2658–2671.

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**:1481–1489.e1410.

Motato Y, de los Santos MJ, Escriba MJ, Ruiz BA, Remohi J, Meseguer M. Morphokinetic analysis and embryonic prediction for blastocyst formation through an integrated time-lapse system. *Fertil Steril* 2016;**105**:376–384.e379.

Paternot G, Wetzels AM, Thonon F, Vansteenbrugge A, Willemen D, Devroe J, Debrock S, D'Hooghe TM, Spiessens C. Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. *Reprod Biol Endocrinol* 2011;**9**: 127.

Petersen BM, Boel M, Montag M, Gardner DK. Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on Day 3. *Hum Reprod* 2016;**31**: 2231–2244.

Polanski LT, Coelho Neto MA, Nastri CO, Navarro PA, Ferriani RA, Raine-Fenning N, Martins WP. Time-lapse embryo imaging for improving reproductive outcomes: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2014;**44**:394–401.

Racowsky C, Kovacs P, Martins WP. A critical appraisal of time-lapse imaging for embryo selection: where are we and where do we need to go? *J Assist Reprod Genet* 2015;**32**:1025–1030.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–428.

Storr A, Venetis CA, Cooke S, Susetio D, Kilani S, Ledger W. Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. *J Assist Reprod Genet* 2015; **32**:1151–1160.

Sundvall L, Ingerslev HJ, Breth Knudsen U, Kirkegaard K. Inter- and intra-observer variability of time-lapse annotations. *Hum Reprod* 2013;**28**: 3215–3221.

WHO. World Health Organization laboratory manual for the examination and processing of human semen. 5th edition. 2010.

Wong C, Chen AA, Behr B, Shen S. Time-lapse microscopy and image analysis in basic and clinical embryo development research. *Reprod Biomed Online* 2013;**26**:120–129.

Wong CC, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, Reijo Pera RA. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol* 2010;**28**:1115–1121.