



www.sciencedirect.com
www.rbmonline.com



ARTICLE

A clinically useful simplified blastocyst grading system




Alison Richardson ^{a,b,*}, Sophie Brearley ^a, Saran Ahitan ^a,
Sarah Chamberlain ^a, Tracey Davey ^a, Lyndsey Zujovic ^a, James Hopkisson ^a,
Bruce Campbell ^b, Nick Raine-Fenning ^{a,b}

^a Nurture Fertility, The East Midlands Fertility Clinic, Nottingham NG10 5QG, UK; ^b Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK

* Corresponding author. E-mail address: alison.richardson@nottingham.ac.uk (A Richardson).



Dr Alison Richardson is a speciality trainee in obstetrics and gynaecology, and is currently taking time out of training and working as a Clinical Research Fellow at Nurture Fertility. She is undertaking a PhD at the University of Nottingham under the supervision of Nick Raine-Fenning and Professor Bruce Campbell. Her special interests include reproductive medicine and early pregnancy development.

Abstract The aim of this study was to investigate whether a new simplified blastocyst grading system (A: fully expanded, clear inner cell mass, cohesive trophectoderm; B: not yet expanded, clear inner cell mass, cohesive trophectoderm; C: small inner cell mass ± irregular trophectoderm ± excluded/degenerate cells) was clinically useful. All day-5 single embryo transfers between 15 June 2009 and 29 June 2012 were reviewed. Implantation, clinical pregnancy and live birth rates were related to embryo quality. Five embryologists were asked to grade and decide the clinical fate of 80 images of day-5 embryos on two occasions 4–6 weeks apart. Implantation, clinical pregnancy and live birth rates decreased with deteriorating embryo quality. A highly significant ($P < 0.01$) difference was observed between the groups. Inter-observer agreement was substantial for grade allocation ($K = 0.63$) and clinical decision-making ($K = 0.66$). Intra-observer agreement ranged from substantial ($K = 0.71$) to almost perfect ($K = 0.88$) for grade allocation, and was almost perfect for clinical fate determination ($K \geq 0.84$). This grading system is quick and easy to use, effectively predicts IVF outcome and has levels of agreement similar to, if not better than, those associated with more complex grading systems. 

© 2015 Reproductive Healthcare Ltd. Published by Elsevier Ltd. All rights reserved.

KEYWORDS: blastocyst, grading, variability

Introduction

Blastocyst culture increases the success rate of assisted reproduction techniques because it permits better embryo

selection after genomic activation, is associated with better endometrial receptivity, or both (Hardarson et al., 2012). Extended culture to the blastocyst stage also enables selection of the most viable embryo from a cohort, thus reducing

the need to transfer multiple embryos to achieve reasonable success rates with a consequent reduction in the incidence of multiple pregnancy (Hardarson et al., 2012).

Time-lapse microscopy, molecular karyotyping and proteomics and metabolomics are increasingly being used in many units to aid identification of the optimum embryo; however, the decision of which blastocyst to transfer is still largely made on the basis of morphological assessments conducted in the IVF laboratory at the time of embryo transfer (Dokras et al., 1993; Gardner et al., 2004). Multiple blastocyst grading systems are in existence to aid this process. Regardless of the grading system used, multiple studies have demonstrated a strong correlation between blastocyst quality and implantation and clinical pregnancy rates (Balaban et al., 2000, 2006; Gardner et al., 2000, 2004).

Most grading systems currently used for assessing the viability of IVF embryos are subjective, relying on visual inspection of morphological characteristics of the embryos that are qualitatively evaluated. Grading based on qualitative criteria is imprecise and inevitably results in inter-observer, and to some extent intra-observer, variability. Gardner et al.'s (1998) grading system, which is used by many clinics, allows for 54 different permutations and hence considerable scope exists for different embryologists to allocate the same blastocyst a different grade (inter-observer variation) or the same embryologist to allocate the same blastocyst a different grade if assessed on a different occasion (intra-observer variation).

Clinically, it is important to try to minimize variability in embryo scoring because the grade of the embryo is used to predict the likelihood of successful treatment, and therefore influences the decision on which embryo to replace and how many embryos to transfer. It also dictates how couples undergoing IVF treatment are counselled about the likelihood of implantation, clinical pregnancy, multiple pregnancy and live birth so that their expectations can be appropriately managed.

It has been suggested that minimizing inter- and intra-observer variability could be achieved by having all the embryo grading done by a single observer or by having multiple embryologists evaluate each embryo and then decide upon the grade by consensus (Baxter Bendus et al., 2006). Neither of these options are practical for most IVF clinics.

Simplifying the grading system will, by definition, reduce variability, as has been shown (Al-Aynati et al., 2003). In 2009, Dr Cecilia Sjoblom devised a simplified blastocyst grading system (Table 1) based on our unit's prior experience of the Gardner system. An essential component of any grading system is that it is not only accurate and reproducible but that it can be used to predict outcome. We therefore investigated our simplified blastocyst grading system to determine whether or not it could be used to predict clinical outcome in terms of implantation, clinical pregnancy and live birth and is consistent and accurate with minimal inter- and intra-observer variability.

Materials and methods

In 2009, when we decided to simplify our blastocyst grading system, it was noted from our experience of using the Gardner grading system that any blastocyst that had a Gardner expansion grade of 4, 5 or 6 (i.e. fully expanded) with an inner

cell mass and trophectoderm grade 'A' or 'B' had the highest implantation rate, so this became the criteria for the grade 'A' blastocysts in our simplified system. Any blastocyst that was not fully expanded but had an inner cell mass and trophectoderm grade 'A' or 'B' had a slightly lower implantation rate, and this became the benchmark for our grade 'B' blastocysts. Any blastocyst that had a Gardner grade 'C' for inner cell mass or trophectoderm, regardless of expansion status, had a much lower implantation rate and this became our grade 'C' blastocyst (Table 1).

Determination of prognostic potential

We reviewed all single (elective or otherwise), fresh or frozen day-5 embryo transfers at Nurture Fertility between 15 June 2009 and 29 June 2012. All participants had undergone IVF-ICSI treatment using a standard long agonist or antagonist protocol, depending on ovarian reserve tests as previously described (Jayaprakasan et al., 2010, 2014).

The following data were collected: age, ethnicity, smoking status, BMI, type of treatment (IVF-ICSI) and grade of blastocyst transferred (as judged by the duty embryologist on the day of embryo transfer). Implantation (defined as a positive urinary pregnancy test performed 18 days after oocyte retrieval), clinical pregnancy (defined as ultrasonographic evidence of at least one fetal heartbeat) and live birth (defined as delivery of a live baby at more than 24 weeks gestation) rates were recorded. Information was obtained from IVF unit records. Embryos that were assigned ambiguous grades (for example A/B, B/C) not compliant with the grading system described above were excluded, as were any transfers where outcome data were not available.

Determination of inter- and intra-observer variability

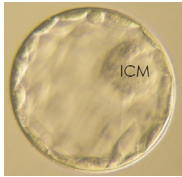

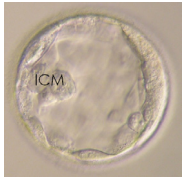


Five embryologists were asked to view 80 still images of day-5 embryos. All embryologists had a life sciences degree and were State Registered Clinical Scientists with the Health and Care Professions Council. They had been qualified for between 7 and 20 years, and had between 2 and 5 years' experience using this simplified blastocyst grading system.

Still images were obtained by one of the authors (SB). The images were randomly grouped into subsets comprising five embryos. Participants were asked to grade the embryos using the simplified blastocyst grading system described above. They were also asked, out of each subset of five images, which embryo they would preferentially transfer if presented with that cohort, which, if any, they would freeze (grade A or B blastocysts only as per unit policy) and which, if any, they would discard. They were blinded to the assessments of the other embryologists to minimize bias.

The same cohort of embryologists were asked to grade the same set of still images 4–6 weeks after the initial assessment. The order of the still images had been randomly manipulated before the second assessment in an attempt to eliminate recall bias.

According to the Medical Research Council Health Research Authority, ethical approval was not required for this

Table 1 The Nurture simplified blastocyst grading system.

Grade	Description	Image
A	Expanded ^a blastocyst, larger in size than a cleavage stage embryo with a thinning zona pellucida. Must have a clear inner cell mass and cohesive trophoctoderm.	
B	Fully formed blastocyst with a clear inner cell mass and cohesive trophoctoderm but not yet expanded with a non-thinning zona pellucida.	
C	Blastocyst with small/no inner cell mass and/or irregular/non-continuous trophoctoderm and/or excluded/degenerate cells.	
Cavitating	Evidence of a cavity of any size forming within the embryo, two distinct cell types not yet visible.	
Compacting	Evidence of any signs of compaction between cells of the embryo.	

^aBlastocoele cavity over 50% of volume (subjective assessment).

study. This was confirmed in April 2015 before submission of the manuscript for consideration for publication.

Statistical analysis

For the determination of prognostic potential, statistical analysis was carried out using SPSS 21 (Statistical Package for Social Sciences; IBM, Chicago, IL, USA). Continuous data were analysed by a Student’s *t*-test or by the Mann–Whitney *U*-test depending on the data distribution. Categorical data were analysed using the chi-squared test, unless more than 20% of the expected values were less than 5, in which case Fisher’s Exact Test was used. Bonferroni corrections were applied for multiple comparisons. When *P* < 0.05, the difference was considered to be statistically significant in all statistical tests.

For the determination of inter- and intra-observer variability, statistical analysis was conducted using Minitab 16 (Minitab Inc). To describe inter- and intra-observer variability, the Fleiss–Kappa statistic and Kendall’s coefficient of

concordance was used. Fleiss–Kappa is a generic term for several similar measures of agreement used with categorical data, which reflect the classification of objects into different groups or categories. Typically, it is used to assess the degree to which two or more raters, examining the same data, agree when it comes to assigning the data into categories. In this study, the Fleiss–Kappa statistic was used to assess the extent by which the embryologists vary in their assignment of a grade of a blastocyst using the simplified blastocyst grading system under evaluation and described above. Complete agreement corresponds to *K* = 1, and lack of agreement corresponds to *K* = 0. A negative value of kappa would mean negative agreement, usually caused by the rater’s tendency toward avoiding a grade assigned to an object by others. Recommended descriptions of numerical *K* values are presented in [Table 2](#) (Landis and Koch, 1977). Similarly, Kendall’s coefficient of concordance can also be used for assessing agreement among raters and ranges from 0 (no agreement) to 1 (complete agreement). Intermediate values of *W* indicate a greater or lesser degree of unanimity among the various responses.

Table 2 Recommended descriptions of numerical K values.

Kappa value	Level of agreement
0.10–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Results

Determination of prognostic potential

Between 15 June 2009 and 29 June 2012, 580 single embryo transfers were carried out on day 5. Thirty-one of these were excluded from further analysis because they were graded as 'A/B' ($n = 9$), 'B/C' ($n = 14$), 'early blast' ($n = 6$), and 'cav/comp' ($n = 2$). A further four were excluded because they were lost to follow-up, and no pregnancy outcome data were available. The remaining 545 single embryo transfers formed the study group: this comprised 225 (41.3%) grade A, 209 (38.3%) grade B and 58 (10.6%) grade C blastocysts and 38 (7.0%) cavitating and 15 (2.8%) compacting embryos.

No significant difference was found in the baseline characteristics of women in each of the five embryo grades in terms of age, body mass index, smoking status, ethnicity and cycle number (Table 3).

Implantation, clinical pregnancy and live birth rates are shown in Figure 1. Implantation, clinical pregnancy and live birth was generally less likely to occur as embryo quality deteriorated. Implantation rates ranged from 79.1% (grade A) to 13.2% (cavitating). Clinical pregnancy and live birth rates ranged from 69.8% (grade A) to 10.5% (cavitating) and 65.8% (grade A) to 6.7% (compacting) respectively. Compacting

embryos were more likely to implant and result in a clinical pregnancy than cavitating embryos but not a live birth. Overall there was a highly significant ($P < 0.01$) difference between the five grades of embryo in all three outcome parameters. This effect was principally due to the highly significant ($P < 0.01$) difference in implantation, clinical pregnancy and live birth rates found between grade A and grade B blastocysts and the significant ($P < 0.05$) difference in clinical pregnancy and live birth rates found between grade B and grade C blastocysts. For implantation, the difference between grade B and grade C blastocysts was not statistically significant but the difference between grade C blastocysts and cavitating embryos was ($P < 0.05$). No significant difference was found in any of the three outcomes among the poorer quality embryos.

Determination of inter-observer variability

Overall, the level of agreement between the five embryologists when assigning a grade to the embryos using the simplified blastocyst grading system was 'substantial' as demonstrated by a kappa score of 0.63 and a Kendall's coefficient value of concordance of 0.89. Agreement was highest for poor quality (compacting embryos, $K = 0.94$; cavitating embryos, $K = 0.79$) and very high quality (grade A blastocysts, $K = 0.71$) embryos (Table 4). Slightly greater variability was observed when assigning blastocysts to grades B ($K = 0.46$) and C ($K = 0.62$), although the level of agreement observed was still 'moderate' or 'substantial' respectively.

Similarly, when determining the fate of the embryos in each cohort, the overall level of agreement was 'substantial' ($K = 0.66$). The level of agreement regarding which, if any, embryos to discard was greater ($K = 0.77$) than the level of agreement regarding which embryo to transfer ($K = 0.64$) and which, if any, to freeze ($K = 0.58$) (Table 5).

Table 3 Baseline characteristics of women according to grade of day-5 embryo transferred.

	A ($n = 225$)	B ($n = 209$)	C ($n = 58$)	Cavitating ($n = 38$)	Compacting ($n = 15$)
Age					
Mean \pm SD	33.24 \pm 4.45	33.21 \pm 4.63	34.22 \pm 5.27	33.44 \pm 4.59	34.16 \pm 4.93
Body Mass Index					
Mean \pm SD	24.02 \pm 3.21	24.50 \pm 3.08	25.06 \pm 3.32	24.68 \pm 3.89	24.48 \pm 4.05
Non-smokers (%)	99.66	96.66	96.55	100	100
Ethnicity (%)					
White	83.11	88.52	94.83	97.37	86.67
Mixed/multiple ethnic groups	0.44	0.00	0.00	0.00	0.00
Asian/Asian British	12.00	10.53	5.17	2.63	6.67
Black/African/Caribbean/Black British	4.00	0.96	0.00	0.00	0.00
Other ethnic group	0.44	0.00	0.00	0.00	6.67
Cycle number (%)					
1	84.44	87.56	81.03	78.95	80.00
2	10.22	6.70	13.79	10.53	13.33
3	1.78	2.87	1.724	2.63	6.67
4	2.67	1.44	3.45	5.26	0.00
5	0.89	1.44	0.00	2.63	0.00

No statistically significant differences were found.

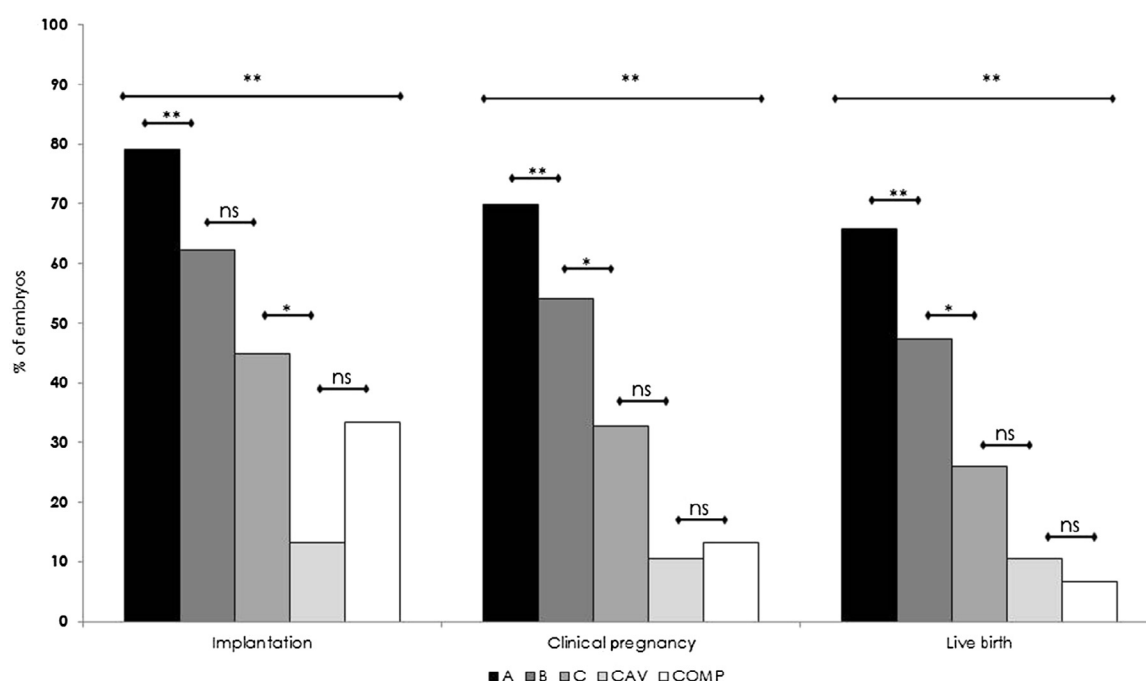


Figure 1 Implantation, clinical pregnancy and live birth rates according to embryo grade on day 5. CAV, cavitating, comp, compacting; NS = not significant. * $P < 0.05$; ** $P < 0.01$.

Table 4 Between operator (inter-observer) consistency of allocation of embryo grade.

Embryo grade	Kappa statistic	Level of agreement
A	0.71	Substantial
B	0.46	Moderate
C	0.62	Substantial
Cavitating	0.79	Substantial
Compacting	0.94	Almost perfect
Overall	0.63	Substantial

Table 5 Between operator (inter-observer) consistency of clinical decision regarding embryo fate.

Decision	Kappa statistic	Level of agreement
Transfer	0.64	Substantial
Freeze	0.58	Moderate
Discard	0.77	Substantial
Overall	0.66	Substantial

Determination of intra-observer variability

Overall, the level of agreement within operators when re-grading the same set of still images ranged from 'substantial' (operator 4, $K = 0.71$) to 'almost perfect' (operator 2, $K = 0.88$) (Table 6). Consistency was greatest when grading the poorest quality embryos as all five operators had an 'almost perfect' level of agreement when classifying an embryo as compacting. Least consistency occurred among all operators when assigning an embryo a grade B but at worst, there

was still a 'moderate' level of agreement (operator 4, $K = 0.57$) and at best an 'almost perfect' level of agreement (operator 2, $K = 0.81$).

Similarly, the level of agreement when re-determining the fate of the embryos in each cohort was 'almost perfect' for all operators with overall K values ranging from 0.84 (operator 3) to 0.94 (operator 2) (Table 7). All operators had 'almost perfect' levels of agreement for all decisions (transfer, freeze and discard) apart from operator 3 who only had a 'substantial' level of agreement when deciding which embryo to transfer as reflected by a K value of 0.75.

Discussion

This study demonstrates both the prognostic potential and the inter- and intra-observer variability of a simplified blastocyst grading system.

The results show that this simplified grading scheme can be used to effectively predict clinical outcome of implantation, clinical pregnancy and live birth with higher quality blastocysts being statistically significantly more likely to yield positive results than poorer quality ones. There is a trend towards compacting embryos being more likely to implant and produce a clinical pregnancy than cavitating embryos but this difference was not statistically significant and not apparent for the more important clinical outcome of live birth. This most likely reflects the small numbers of embryos in these two groups.

Several studies (Arce et al., 2006; Baxter Bendus et al., 2006; Matson, 1998; Paternot et al., 2009; Ruiz de Assin et al., 2009) have analysed the inter- and intra-observer variability associated with grading cleavage stage embryos with mixed results. Few studies, however, have reported on the inter- and

Table 6 Within operator (intra-observer) consistency of allocation of embryo grade.

Embryo Grade	Kappa statistic				
	Operator 1	Operator 2	Operator 3	Operator 4	Operator 5
A	0.82	0.84	0.84	0.78	0.87
B	0.75	0.81	0.77	0.57	0.77
C	0.82	0.96	0.92	0.70	0.89
Cavitating	0.88	1.0	0.88	0.69	1.0
Compacting	0.85	1.0	0.85	1.0	1.0
Overall	0.79	0.88	0.84	0.71	0.86

Table 7 Within operator (intra-observer) consistency of clinical decision regarding embryo fate.

Decision	Kappa statistic				
	Operator 1	Operator 2	Operator 3	Operator 4	Operator 5
Transfer	0.84	0.92	0.75	1.0	0.84
Freeze	0.87	0.92	0.80	0.87	0.82
Discard	0.97	0.97	0.94	0.87	0.91
Overall	0.90	0.94	0.84	0.90	0.86

intra-observer variability encountered when grading blastocysts using the various different grading systems. The variation between two observers can introduce bias and make it difficult to interpret results (Sundvall et al., 2013); however, our results show that between embryologists, there is overall a 'substantial' level of agreement when grading embryos using the simplified blastocyst grading system. Inter-observer variation was lowest when assessing very high or very poor quality embryos and highest when assessing embryos in the middle of the spectrum. A 'substantial' level of agreement was also found between embryologists when deciding upon the clinical fate of the embryos in each cohort presented and this is consistent with the levels of agreement observed during the grade allocation. The variation within an observer is assumed to be random and, as such, does not cause bias in itself but affects precision. Knowledge about precision is crucial when assessing the relevance of embryo selection parameters (Sundvall et al., 2013). Our data show that the intra-observer variability associated with the simplified blastocyst grading system is minimal. Overall, all operators exhibited 'substantial' or 'almost perfect' levels of agreement within themselves on the two occasions that they were asked to grade and decide upon the clinical fate of the same set of 80 still images. This suggests that our simplified blastocyst grading system is precise.

Some might argue that our simplified grading system is similar to the one proposed by Dokras et al. (1993), and indeed it does share some similarities. Our grade A blastocyst is similar to Dokras' grade 1 blastocyst as both are defined by an expanded cavity with a distinct inner cell mass region and trophectoderm layer. Similarly our grade B and Dokras' grade 2 blastocysts may both be considered to be 'slow/late developers' and our grade C and Dokras' grade 3 blastocysts may both contain degenerate cells. Our grade B blastocysts, unlike Dokras' grade 2 blastocysts, however, do not contain vacuoles and are more likely to resemble our grade A blastocysts

within a period of hours rather than a period of days. They have a distinct inner cell mass and cohesive trophectoderm. The only difference between a grade B and a grade A blastocyst according to our grading system is the degree of expansion of the blastocoel cavity. In contrast, the difference between a grade 1 and grade 2 blastocyst according to Dokras is rather more striking. Furthermore, Dokras' definition of a grade 3 blastocyst only makes reference to the presence of degenerative foci in the inner cell mass. Our grade C blastocysts are mainly defined by the presence of a small or absent inner cell mass, an irregular or non-continuous trophectoderm with or without excluded or degenerate cells, or both. So, although some similarities exist between our grading system and Dokras', we believe ours to be more specific, taking into account all the morphological qualities seen on assessment of day 5 embryos, including the degree of expansion of the blastocoel cavity and the quality of both the inner cell mass and trophectoderm (Cutting et al., 2008; Hill et al., 2013). The fact that the grading system is still simplistic should hopefully ensure that it can be easily adopted by other embryologists in other IVF units. This claim would however need to be validated prior to widespread implementation.

Our study is strengthened by the fact that 545 single embryo transfers were included in the analysis of prognostic potential. Unfortunately, the numbers included in the cavitating and compacting groups were limited. This is difficult to resolve as double embryo transfer is often recommended under such circumstances in order to optimize pregnancy rates. An additional strength is that we have reported live birth and not just implantation and clinical pregnancy rates.

For the determination of inter- and intra-observer variability, each embryologist was asked to grade a large number of day 5 embryos and to record their clinical decision making as a consequence of their grade allocation. In part, this was an attempt at quantifying whether the use of a prognostic model truly improves the user's decision making and ultimately

patient outcome which is an important component of prognostic research (Moons et al., 2009). Although only five embryologists were included in the study, all had a similar experience and background, which has been shown to be a successful method to limit bias (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2011; Ruiz de Assin et al., 2009). One weakness is that still images were used for the assessment of quality, which provides a somewhat artificial environment for the embryologists as they were unable to manipulate the embryos as they would in routine clinical practice in order to gain a three-dimensional view. Similarly still images only afford the operator a snapshot view of each embryo and as such, any embryos caught in the processes of cytokinesis or cellular reorganization for example during the 30 second window of its development in which it was photographed may be misclassified. Since all embryologists, were exposed to the same restraints, any disadvantage created by these limitations was universally experienced by all and should not have significantly affected the inter- and intra-observer variability, which is principally what was being assessed.

It is necessary to validate the grading system, both internally and externally, before widespread implementation (Altman et al., 2009). Additionally, many studies have demonstrated that stimulation protocols are related to endometrial receptivity (Devroey et al., 2004; Valbuena et al., 1999); therefore, it would be very interesting for futures studies to do the same analysis conducted here, but in frozen cycles, thereby reducing the endometrial receptivity variant. This may, however, prove difficult practically as many IVF units (including ours) only freeze good-quality embryos and less frequently transfer single embryos when frozen. It would also be interesting, based on the fact that pregnancies occurring as a consequence of IVF are associated with an increased incidence of several obstetric and perinatal complications, to investigate whether, once implanted, all embryos have a similar potential or whether 'good' embryos result in 'good' pregnancies and healthy babies compared with poorer quality embryos. Such knowledge would be useful in clinical practice and the adoption of elective single embryo transfers makes analysis of these effects possible. Some evidence shows that poor-embryo quality is not associated with adverse obstetric or perinatal outcomes (Oron et al., 2014); however, data are limited and further research in this area is warranted.

In conclusion, this study demonstrates the prognostic potential and inter- and intra-observer variability of our simplified blastocyst grading system. The grading scheme was able to effectively predict clinical outcome in terms of implantation, clinical pregnancy and live birth. Slight variation existed both between and within embryologists grading the embryos but, overall, levels of agreement were similar to, if not better than, those associated with more complex grading systems. This, combined with the fact it is quick and easy to use at the time of embryo transfer, make our simplified blastocyst grading system clinically very useful.

Acknowledgements

The authors would like to thank their former colleague, Dr Cecilia Sjoblom for her contribution towards the development

of this simplified blastocyst grading system. The study was funded by Nurture Fertility and the University Of Nottingham.

References

- Al-Aynati, M., Chen, V., Salama, S., Shuhaibar, H., Treleaven, D., Vincic, L., 2003. Interobserver and intraobserver variability using the Fuhrman grading system for renal cell carcinoma. *Arch. Pathol. Lab. Med.* 127, 593–596.
- Altman, D.G., Vergouwe, Y., Royston, P., Moons, K.G., 2009. Prognosis and prognostic research: validating a prognostic model. *BMJ* 338, b605.
- Arce, J.C., Ziebe, S., Lundin, K., Janssens, R., Helmggaard, L., Sorensen, P., 2006. Interobserver agreement and intraobserver reproducibility of embryo quality assessments. *Hum. Reprod.* 21, 2141–2148.
- Balaban, B., Urman, B., Sertac, A., Alatas, C., Aksoy, S., Mercan, R., 2000. Blastocyst quality affects the success of blastocyst-stage embryo transfer. *Fertil. Steril.* 74, 282–287.
- Balaban, B., Yakin, K., Urman, B., 2006. Randomized comparison of two different blastocyst grading systems. *Fertil. Steril.* 85, 559–563.
- Baxter Bendus, A.E., Mayer, J.F., Shipley, S.K., Catherino, W.H., 2006. Interobserver and intraobserver variation in day 3 embryo grading. *Fertil. Steril.* 86, 1608–1615.
- Cutting, R., Morroll, D., Roberts, S.A., Pickering, S., Rutherford, A., 2008. Elective single embryo transfer: guidelines for practice British Fertility Society and Association of Clinical Embryologists. *Hum. Fertil. (Camb.)* 11, 131–146.
- Devroey, P., Bourgain, C., MacKlon, N.S., Fauser, B.C., 2004. Reproductive biology and IVF: ovarian stimulation and endometrial receptivity. *Trends Endocrinol. Metab.* 15, 84–90.
- Dokras, A., Sargent, I.L., Barlow, D.H., 1993. Human blastocyst grading: an indicator of developmental potential? *Hum. Reprod.* 8, 2119–2127.
- Gardner, D.K., Schoolcraft, W.B., Wagley, L., Schlenker, T., Stevens, J., Hesla, J., 1998. A prospective randomized trial of blastocyst culture and transfer in in-vitro fertilization. *Hum. Reprod.* 13, 3434–3440.
- Gardner, D.K., Lane, M., Stevens, J., Schlenker, T., Schoolcraft, W.B., 2000. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil. Steril.* 73, 1155–1158.
- Gardner, D.K., Surrey, E., Minjarez, D., Leitz, A., Stevens, J., Schoolcraft, W.B., 2004. Single blastocyst transfer: a prospective randomized trial. *Fertil. Steril.* 81, 551–555.
- Hardarson, T., Van Landuyt, L., Jones, G., 2012. The blastocyst. *Hum. Reprod.* 27 (Suppl. 1), i72–i91.
- Hill, M.J., Richter, K.S., Heitmann, R.J., Graham, J.R., Tucker, M.J., Decherney, A.H., Browne, P.E., Levens, E.D., 2013. Trophoctoderm grade predicts outcomes of single-blastocyst transfers. *Fertil. Steril.* 99, 1283–1289, e1.
- Jayaprakasan, K., Hopkisson, J., Campbell, B., Johnson, I., Thornton, J., Raine-Fenning, N., 2010. A randomised controlled trial of 300 versus 225 IU recombinant FSH for ovarian stimulation in predicted normal responders by antral follicle count. *BJOG* 117, 853–862.
- Jayaprakasan, K., Pandian, D., Hopkisson, J., Campbell, B.K., Maalouf, W.E., 2014. Effect of ethnicity on live birth rates after in vitro fertilisation or intracytoplasmic sperm injection treatment. *BJOG* 121, 300–306.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

- Matson, P.L., 1998. Internal quality control and external quality assurance in the IVF laboratory. *Hum. Reprod.* 13 (Suppl. 4), 156–165.
- Moons, K.G., Royston, P., Vergouwe, Y., Grobbee, D.E., Altman, D.G., 2009. Prognosis and prognostic research: what, why, and how? *BMJ* 338, b375.
- Oron, G., Son, W.Y., Buckett, W., Tulandi, T., Holzer, H., 2014. The association between embryo quality and perinatal outcome of singletons born after single embryo transfers: a pilot study. *Hum. Reprod.* 29, 1444–1451.
- Paternot, G., Devroe, J., Debrock, S., D’Hooghe, T.M., Spiessens, C., 2009. Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. *Reprod. Biol. Endocrinol.* 7, 105.
- Paternot, G., Wetzels, A.M., Thonon, F., Vansteenbrugge, A., Willemen, D., Devroe, J., Debrock, S., D’Hooghe, T.M., Spiessens, C., 2011. Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. *Reprod. Biol. Endocrinol.* 9, 127.
- Ruiz de Assin, R., Clavero, A., Gonzalvo, M.C., Ramirez, J.P., Zamora, S., Fernandez, A., Martinez, L., Castilla, J.A., 2009. Comparison of methods to determine the assigned value in an external quality control programme for embryo evaluation. *Reprod. Biomed. Online* 19, 824–829.
- Sundvall, L., Ingerslev, H.J., Breth Knudsen, U., Kirkegaard, K., 2013. Inter- and intra-observer variability of time-lapse annotations. *Hum. Reprod.* 28, 3215–3221.
- Valbuena, D., Jasper, M., Remohi, J., Pellicer, A., Simon, C., 1999. Ovarian stimulation and endometrial receptivity. *Hum. Reprod.* 14 (Suppl. 2), 107–111.
- Declaration: The authors report no financial or commercial conflicts of interest.*

Received 8 April 2015; refereed 24 June 2015; accepted 30 June 2015.