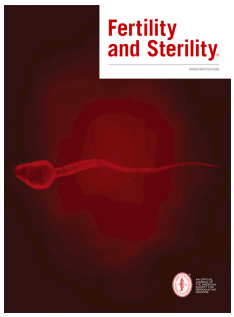
Journal Pre-proof



External validation of a fully automated evaluation tool: a retrospective analysis on 68,471 scored embryos.

Lorena Bori, PhD, Marco Toschi, MSc, Rebeca Esteve, MSc, Arantza Delgado, PhD, Antonio Pellicer, M.D, Marcos Meseguer, PhD

PII: S0015-0282(24)02300-8

DOI: https://doi.org/10.1016/j.fertnstert.2024.10.006

Reference: FNS 35171

To appear in: *Fertility and Sterility*

Received Date: 6 March 2024

Revised Date: 4 October 2024

Accepted Date: 7 October 2024

Please cite this article as: Bori L, Toschi M, Esteve R, Delgado A, Pellicer A, Meseguer M, External validation of a fully automated evaluation tool: a retrospective analysis on 68,471 scored embryos., *Fertility and Sterility* (2024), doi: https://doi.org/10.1016/j.fertnstert.2024.10.006.

Running Title: Fully automated embryo evaluation model

Article Title: External validation of a fully automated evaluation tool: a retrospective analysis on 68,471 scored embryos.

Lorena Bori, PhD[a]; Marco Toschi, MSc[b]; Rebeca Esteve, MSc[a]; Arantza Delgado, PhD[a]; Antonio Pellicer, M.D[b]; Marcos Meseguer, PhD[a,]

[a] IVIRMA Global Research Alliance, IVIRMA Valencia, IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Spain.

[b] IVIRMA Global Research Alliance, IVIRMA Roma, Italy.

Corresponding Author:

Author Name: Lorena Bori

Academic Affiliation: IVIRMA Global Research Alliance, IVIRMA Valencia, IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Spain.

Mailing Address: 8, Laurona Street, Lliria, Valencia 46160, Spain.

Phone: +34 610 55 37 99

Email: Lorena.bori@ivirma.com

Article type: Cohort study

Funding Statement: ISCII (PI21/00283), cofounded by ERDF, "A way to make Europe" awarded to Marcos Meseguer.

Disclosure Statement: None

Attestation Statement:

• Data regarding any of the subjects in the study has not been previously published unless specified.

• Data will be made available to the editors of the journal for review or query upon request.

Data Sharing Statement: The data underlying this article will be shared on reasonable request to the corresponding author.

Word Count: 349 words for abstract and 4,010 words for text

Capsule: Fully automated embryo score effectively ranked embryos based on their potential to implant and result in a live birth.

Structured Abstract

Objective: To externally validate a fully automated embryo classification in in vitro fertilization (IVF) treatments.

Design: Retrospective cohort study

Subjects: A total of 6,434 patients undergoing 7,352 IVF treatments contributed 70,456 embryos.

Exposure: Embryos were evaluated by conventional morphology and retrospectively scored using a fully automated deep learning-based algorithm across conventional IVF, oocyte donation, and PGT-A cycles.

Main Outcome Measures: The primary outcomes were implantation and live birth including odds ratios (ORs) from generalized estimating equation (GEE) models. Secondary outcomes were embryo morphology, euploidy and miscarriage. Exploratory outcomes included comparison between conventional methodology and artificial intelligence (AI) algorithm with areas under the ROC curves (AUCs), agreement degree between AI and embryologists, Cohen's Kappa coefficient and relative risk (RR).

Results: Implantation and live birth rates increased as the automatic embryo score rose. The GEE model, controlling for confounders, showed the automatic score was associated with an OR of 1.31 (95%CI[1.25-1.36]) for implantation in treatments using oocytes from patients, and an OR of 1.17 (95%CI[1.14-1.20]) in the oocyte donation program, with no significant association in PGT-A treatments. For live birth, the ORs were 1.27 (95%CI[1.21-1.33]) for patients, 1.16 (95%CI[1.13-1.19]) for donors, and 1.05 (95%CI[1-1.10]) for PGT-A cycles. The average score was higher in embryos with better morphology, in euploid embryos compared to aneuploid embryos, and in embryos that resulted in a full-term pregnancy compared to those that miscarried. Concordance between the highest-scoring embryo and the embryo with the best conventional morphology was 71.4%(95%CI[67.7%-75.0%]) in treatments with patient oocytes and 61.0%(95%CI[58.6%-63.4%]) in the oocyte donation program. Overall, the Cohen's Kappa coefficient was 0.63. The automatic embryo score showed similar AUCs to conventional morphology, although implantation was higher when the transferred embryo matched the highest-scoring embryo from each cohort (57.36% vs. 49.98%). RR indicated a 1.14-fold increase in implantation likelihood when the top-ranked embryo was transferred.

Conclusion: Fully automated embryo scoring effectively ranked embryos based on their potential for implantation and live birth. The performance of the conventional methodology was comparable to that of the artificial intelligence-based technology; however, better clinical outcomes were observed when the highest-scoring embryo in the cohort was transferred.

Keywords: Automation, embryo selection, implantation, live birth, concordance.

INTRODUCTION

Artificial intelligence (AI) pervades our daily lives, impacting everything from smartphone applications to medical diagnostics. When properly trained on extensive and reliable data, AI significantly enhances convenience and healthcare quality. Its application in human assisted reproductive technology (ART) for diagnostics is growing rapidly (1).

AI holds great potential in improving in vitro fertilization (IVF) processes. One of its most prominent uses is in embryo selection, where embryos from a cohort must be prioritized for transfer or cryopreservation based on their likelihood of successful implantation. Currently, blastocyst grading is subjective and time-consuming. A notable advantage of AI-based methods is their objectivity, consistency, standardization, and efficiency in embryo evaluation (1). Considerable research has assessed AI's effectiveness in predicting pregnancy outcomes following blastocyst transfer (2–4).

There are variations in how different AI models approach automation. Some methods rely on manual preselection by embryologists, categorizing them as semi-automated. These models are often trained on transferred embryos, limiting their exposure to poor-quality embryos (5–7). In contrast, other methods aim for full automation, regardless of whether they were transferred or not (8,9).

The iDAScore® model, developed by Vitrolife in Aarthus, Denmark, is a commercial deep-learning model designed to assess blastocyst quality from time-lapse videos. The exact variables used by these models remain unknown, as deep-learning techniques do not require predefined feature extraction. Nonetheless, it has been reported a significant correlation between iDAScore version 1.0 (v1) and both morphokinetic and morphological characteristics of pre-implantation embryos (10). For example, factors like cell number, blastomere fragmentation, and blastocyst morphology significantly correlate with this automated scoring system. Embryos with lower scores often exhibit delayed compaction, blastulation, and blastocyst expansion. Several studies have demonstrated a significant correlation between iDAScore v1 and various reproductive outcomes, such us implantation, ongoing pregnancy, miscarriage, live birth, and euploid rates (2–4,11). Recently, this model was updated to version 2 (iDAScore v2), which retains the same deep-learning architecture but incorporates a larger training dataset and the ability to provide scores on day 2/3. It uses 3D convolutions, which allow for the simultaneous identification of spatial (morphological) and temporal (morphokinetic) patterns in time-lapse data (12). According to the developers, this model was calibrated to establish a linear relationship between its predictions and implantation rates. It was trained on a dataset of 181,428 embryos from 22

IVF clinics worldwide. The automatic score was internally validated on embryos transferred after 2, 3 and 5+ days of incubation (12). However, no published validations exist from independent clinics that were not involved in the model's development, which assess the algorithm's performance under real-world conditions across diverse populations. The generalization of algorithms based on morphokinetic parameters has been challenging, as they often fail to perform well in independent laboratories (13,14). Conducting external validation is essential to confirm the reliability, effectiveness, and applicability of the embryo selection algorithm in various clinical settings.

The primary aim of this study was to externally validate the iDAScore v2 algorithm's ability to rank blastocysts based on their potential for implantation and live birth. Secondarily, we analyzed its relationship with embryo morphology, euploidy rate, and miscarriage. Additionally, we compared the AI-based scoring with conventional embryo evaluation and selection methods.

MATERIAL AND METHODS

Study Population

This retrospective analysis included a total of 70,456 embryos cultured at IVI Valencia (Spain) over five years, from 2016 to 2021. The study focused on patients and recipients undergoing ICSI cycles, whose embryos were cultured until the fifth or sixth day of development using the time-lapse systems EmbryoScope and EmbryoScope Plus (Vitrolife, Aarthus, Denmark). Embryos excluded from the analysis were those that could not be scored by the iDAScore v2.0 algorithm (e.g., bubble formation or loss of embryo development recording).

Table 1 shows the composition of the study population. Out of the total population, 68,471 embryos were assessed by the automatic score and 7,082 embryos were included in the preimplantation genetic testing for aneuploidy (PGT-A) program. Ultimately, 10,054 known implantation data (KID) embryos resulted from 9,116 blastocyst transfers.

Ovarian Stimulation and Uterine Receptivity

Donors underwent controlled ovarian stimulation with a GnRH agonist protocol using Decapeptyl 1 (Ipsen Pharma, Spain) until at least eight follicles reached an average diameter of 18 mm, followed by transvaginal oocyte retrieval 36 hours later. Recipients received endometrial preparation as described by Cerrillo et al. (2017)(15), followed by 400 mg of vaginal micronized progesterone (Progeffik, Lab. Effik, Madrid, Spain) every 12 hours after embryo transfer for luteal phase support.

Ovarian stimulation for patients undergoing IVF with autologous oocytes used GnRH antagonist protocols, occasionally including long GnRH agonist protocols or LH surge suppression (16). Gonadotropin doses (recombinant FSH ranging from 150 to 300 IU/day) were adjusted based on patient characteristics, often combined with 75 IU of highly purified HMG (Menopur, Ferring Pharmaceuticals). Estrogen was administered from the mid-luteal phase in GnRH antagonist cycles. Long GnRH agonist cycles involved administering 0.1 mg/day of GnRHa (Decapeptyl, Ipsen Pharma) from day 21 of the previous cycle.

Transvaginal ultrasounds were performed on stimulation Day 5 and every two days afterward. GnRH antagonist (Orgalutran, MSD; or Cetrotide, Merck Serono) was introduced when at least one follicle reached an average diameter of 14 mm. Final oocyte maturation was triggered using 0.2 mg GnRHa and/or rhCG 250 mcg (Ovitrelle, Merck Serono) when at least three follicles reached an average diameter of 17-18 mm in GnRH antagonist cycles. In long GnRH agonist cycles, only rhCG was administered.

Oocyte Retrieval and ICSI

Follicular aspiration was performed via transvaginal ovarian puncture to retrieve oocytes, which were then washed in Gamete medium (Cook Medical, Australia) and cultured in Origio fertilization medium (Cooper Surgical, Denmark) at 6% CO2, 5% O2, and 37°C. Oocyte denudation, 4 hours post-retrieval, involved mechanical and chemical methods (pipetting with 40 IU/ml hyaluronidase). Intracytoplasmic sperm injection (ICSI) was conducted using Origio fertilization medium at 400x magnification with an Olympus IX7 microscope.

After ICSI, oocytes were placed in pre-equilibrated EmbryoSlides (Vitrolife) until they reached the blastocyst stage. The culture medium (Gems, Genea Biomedx, Sydney, Australia) volume was 28 µl for conventional EmbryoScope or 180 µl for EmbryoScope Plus, with embryos covered in 1.6 ml of mineral oil during culture.

Embryo incubation, scoring and selection

Embryos were cultured individually in the conventional EmbryoScope or in groups of up to eight in the EmbryoScope Plus system until the fifth/sixth day of development. Automatic imaging was performed every 10-20 minutes, capturing images from up to 11 focal planes. External computer analysis software, specifically the EmbryoViewer™ workstation by Vitrolife, was used to assess embryo development.

Fertilization evaluation was conducted 16 to 19 hours after ICSI and confirmed by the presence of two pronuclei and two polar bodies. The division times to 2 cells (t2), 3 cells (t3), 4 cells (t4),

5 cells (t5), as well as the time to blastocyst formation (tB), and the quality of the inner cell mass (ICM) and trophectoderm, were automatically recorded using the guided annotations tool within the EmbryoViewer. Senior embryologists evaluated blastocysts at 114-118 hours (Day 5) and 136-140 hours (Day 6) post-insemination. The assessment focused on blastocoel expansion and the quality of the ICM and trophectoderm, following the criteria set by the Association for the Study of the Biology of Reproduction (ASEBIR) (refer to Supplementary Tables S1, S2, S3, and S4). Additionally, embryos were retrospectively scored by the iDAScore v2 algorithm with a numerical score from 1 (lowest) to 9.9 (highest), indicating the likelihood of implantation.

PGT-A

In cycles involving PGT-A, embryos underwent assisted hatching on Day 3 following cell counting, using the Hamilton-Thorne LykosVR laser. The time-lapse video paused during the procedure and resumed when the plate was reinserted, with automatic refocusing. Once the embryos reached the blastocyst stage, 5–6 trophectodermal cells were biopsied and their ploidy was assessed using Next Generation Sequence (NGS) technology from Thermo Fisher Scientific, USA.

Embryo transfer and luteal support

The best morphologically graded embryo of each cohort (89.7%) or the two best embryos were selected for transfer following ASEBIR guidelines. Subsequent frozen-thawed embryo transfers were conducted upon the patient's request, depending on availability. Luteal support involved administrating daily doses of micronized intravaginal progesterone (400 mg after fresh transfers and 800 mg following frozen-thawed transfers) (Progeffik, Effik, Spain). Biochemical pregnancies were determined through blood βhCG levels, measured 11 days after embryo transfer, while implantation was confirmed at the eighth week of pregnancy via ultrasound observation of the gestational sac. Live births were confirmed through direct communication with the patients.

Outcome measures and statistical analysis

Primary outcome: Contribution of the automatic score to implantation and live birth rates.

Embryo scores were compared within groups based on implantation and live birth outcomes (positive versus negative). Statistical analysis was performed using ANOVA and chi-squared tests to compare scores.

The performance of the iDAScore ranking in predicting implantation and live birth was assessed using generalized estimating equation (GEE), measuring the correlation between the classification system and the outcome while adjusting for potential confounding factors. This

approach treated multiple embryos from a single patient as an intra-subject variable, controlling for patient-related confounders. The final GEE model included only the combination of variables that produced the most optimal model. Finally, a GEE model was created with automatic score as predictive variable and the following variables as potential confounders: oocyte origin (donated versus autologous), type of embryo transfer (fresh versus frozen), oocyte age, patient BMI, day of embryo transfer (fifth versus sixth day of development), culture strategy (group versus single), blastocyst morphology (A versus C grade and B versus C grade), and indication for infertility treatment (e.g., female etiology, male etiology, mixed etiology, social or unknown factor). Additional GEE models were constructed independently for conventional cycles with autologous oocytes across different age groups, the oocyte donation program, and PGT-A cycles.

Odds ratios (ORs) indicating the impact of variables included in the GEEs on the outcome measures were presented with 95% confidence intervals (CIs), and statistical significance was determined for P-values <0.05. To account for potential effects of unmeasured variables, E-values for each OR were provided, calculated using an online E-value calculator (17,18). These provided insight into the minimum strength of association an unmeasured confounder would require to explain a particular association between cycles and outcomes.

Secondary outcomes: Relationship between automatic score and embryo morphology, euploidy rate, and miscarriage.

Embryo scores were compared within groups based on morphological grade (A,B,C,D and excluded), ploidy status (euploid vs. aneuploid), and miscarriage (positive vs. negative). Statistical analysis involved cross-tables, ANOVA and chi-squared tests

Exploratory outcomes: Comparison between the automatic score and conventional morphology-based methodology.

Similar GEEs were formulated using ASEBIR grading as a predictor instead of iDAScore, with analyses conducted for different patient age groups (<35 y.o., 35-37 y.o. , 38-40 y.o. and >40 y.o.), the oocyte donation program and PGT-A treatments. Receiver operating characteristic (ROC) curves were generated from the GEE models' probability values, with the areas under the ROC curves (AUC) calculated as evaluation metrics for model performance.

The median iDAScore for the set of transferred embryos was calculated. This allowed for the subdivision of morphological categories into A+, B+, C+ (automated score above the median) and A-, B-, C- (automated score below the median). Implantation rates were then compared

across each subcategory using ANOVA and chi-squared tests. Finally, GEEs were performed with iDAScore as a predictor within each morphological category.

Concordance between the embryologist's decision and the model's decision (i.e. the highest scoring embryo according to the iDAScore within each cohort) in the first transfer was measured as a simple overall agreement percentage (cases where both evaluators agreed divided by the total number of cases). Cohen's Kappa coefficient was calculated to adjust for chance agreement. Kappa values were interpreted as follows: very good (0.81–1.00), good (0.61–0.80), moderate (0.41–0.60), fair (0.21–0.40), and poor (≤0.20) (19). Finally, relative risk (or incidence of implantation) was calculated as the ratio of the probability of implantation when the highest-scoring embryo was transferred versus when it was not.

Ethical Approval

The procedure and protocol were approved by an institutional review board (IRB reference: 1902-VLC-018-MM) that regulates and approves database analysis and clinical IVF procedures for research at IVI Valencia. In addition, the project complies with the Spanish law governing assisted reproductive technologies (14/2006).

RESULTS

Out of the total population, 10,054 were known implantation data (KID) embryos, including both fresh and frozen transfers. The clinical outcomes for these embryos were as follows: a biochemical pregnancy rate of 59.5%, an implantation rate of 50.5%, and a live birth rate of 42.3%.

Contribution of the automatic score to implantation and live birth rates.

The distribution of automatically scored embryos based on treatment success (implantation and live birth) can be visualized in Supplementary Fig. S1. Positive implantation embryos scored 6.8±2.2, while negative implantation embryos scored 5.7±2.5 ($P<0.001$). Similarly, embryos resulting in a live birth had an average iDAScore of 6.9 ± 2.2, compared to 5.9 ± 2.5 for those that did not result in a live birth ($P < 0.001$). When the embryos were divided into quartiles of similar sample sizes, a significant trend was observed ($P<0.001$), with better clinical outcomes associated with higher iDAScores (Fig. 1).

The iDAScore was associated with higher odds of both implantation and live birth. The generalized estimating equation (GEE) model indicated that each one-unit increase in iDAScore was associated with an OR of 1.18 for implantation (95% CI [1.15-1.22], $P < 0.001$). Similarly,

each one-unit increase in iDAScore was associated with an OR of 1.17 for live birth (95% CI [1.13-1.20], P < 0.001). The model's performance was comparable for both fresh and frozen embryo transfers. For frozen embryo transfers, each unit increase in iDAScore was associated with an OR of 1.20 for implantation (95% CI [1.17-1.24], AUC = 0.643, P < 0.001), while the OR for live birth was 1.18 (95% CI [1.14-1.21], AUC = 0.635, P < 0.001). For fresh embryo transfers, the OR for implantation was 1.22 (95% CI [1.18-1.26], AUC = 0.647, P < 0.001), and the OR for live birth was 1.22 (95% CI [1.17-1.26], AUC = 0.637, P < 0.001). The embryo scores significantly contributed to the implantation outcome in conventional treatments and oocyte donation program, but not in PGT-A cycles. However, they did contribute to the prediction of live birth across all subpopulations separately, as presented in Table 2.

Relationship between automatic score, embryo morphology, euploidy rate, and miscarriage.

Embryo scores varied significantly across different ASEBIR morphological categories (Supplementary Fig. S2). A-grade embryos had an average iDAScore of 8.0 ± 1.4 (n=2,775), B-grade embryos averaged 6.4 ± 2.2 (n=15,523), C-grade embryos averaged 3.8 ± 2.3 (n=11,047), D-grade embryos averaged 1.7 ± 1.3 (n=12,778), and excluded embryos had an average score of 1.2 ± 0.7 (n=26,348) (P < 0.001). Euploid embryos had a higher mean score compared to aneuploid embryos (6.3±2.5; n=3,208 vs. 5.1±2.6; n=3,874; p<0.001) (Supplementary Fig. S3), and euploidy rate increased as the embryo score was higher (Supplementary Fig. S4).

In cycles that did not involve PGT-A, embryos that resulted in miscarriage had a lower average score compared to those that led to delivery (6.6±2.3 vs. 6.9±2.2; p<0.001) (Supplementary Fig. S5). However, a consistent linear decrease in miscarriage rates was not observed as the iDAScore increased (Supplementary Fig. S6).

Comparison between the automatic score and conventional morphology-based methodology.

The comparison between the conventional methodology and the automatic scoring in terms of OR and AUC is summarized in Table 3 for patients from different age groups, for the oocyte donation program and for PGT-A treatments.

Embryos selected by senior embryologists for single embryo transfer in the fresh cycle matched the top-ranked embryo on day 5 in 62.68% of cases (95% CI [60.85%-64.51%], n=2,675). Cohen's Kappa coefficient was 0.63. The mean difference between the iDAScore of the selected embryo and the highest-scoring embryo for non-matching cases was 1.23±1.15. In conventional treatments using autologous oocytes, concordance was higher (71.38%, 95% CI [67.72%-75.04%) than in oocyte donation programs (60.97%, 95% CI [58.57%-63,37%]).

In cycles where Day 5 fresh embryo transfer did not result in implantation (n=1,086), there was a 36.25% rate of discordance 95% CI [33.39%-39.11%], between the embryologists' choice and the iDAScore system. Among these discordant cases, 83.69% had good quality embryos (A or B grade) available with higher scores. When multiple good quality embryos were available, the implantation rate was similar in concordant and discordant cases (63.72% vs. 60.29%; p=0.164). However, for A+B embryos (n=7,806), implanted embryos had a significantly higher iDAScore compared to non-implanted embryos (7.3 ± 1.8 vs. 6.6 ± 2.1, P < 0.001). Implantation rates for each morphological subcategory according to the automatic scoring are shown in Supplementary Fig. S7.

The performance of iDAScore within each morphological quality group showed significant associations. For A-grade embryos (n=1,775), each one-unit increase in iDAScore was associated with an OR of 1.22 (95% CI [1.11–1.33], P < 0.001). For B-grade embryos (n=6,031), the OR was 1.12 (95% CI [1.09–1.16], P < 0.001), and for C-grade embryos (n=2,248), the OR was 1.15 (95% CI [1.10–1.20], P < 0.001).

In general, for fresh and frozen Day 5 single embryo transfers (n=7,126), implantation rates were higher when the transferred embryo matched the highest-scoring embryo: 57.36% in matches vs. 49.98% in mismatches (p<0.001). The RR showed that patients were 1.14 times more likely to achieve pregnancy if the transferred embryo was the one with the highest score assigned by iDAScore v2.

DISCUSSION

Our external validation confirms that the iDAScore v2 algorithm effectively ranks blastocysts according to their potential for implantation and live birth. Moreover, higher embryo scores were associated with better morphological quality, euploid embryos, and a reduced probability of miscarriage after transfer. While the performance of the artificial intelligence based model did not differ substantially from the performance of the conventional methodology for embryo evaluation, it would be useful for further classifying embryos of similar morphological categories to select a single embryo with the highest livebirth potential in a cohort.

The scientific community is already aware of automated scoring assigned by various algorithms is linked to clinical outcomes (3,24–28). Specifically, Ueno 2023 (29) is the only study that has assessed iDAScore v2 for blastocyst transfers, reporting an AUC of 0.736 for predicting ongoing pregnancy. However, the generalizability of those results is questionable, as the study population was limited to patients undergoing minimal stimulation and natural IVF cycles, and

the clinic involved was also part of the algorithm's development process. Although our findings align with theirs, our AUCs were lower across most treatment scenarios (Table 3). This could be attributed to our study being a fully independent external validation, which provides a more realistic assessment of model performance in clinical practice. Moreover, conducting comparisons proves 'challenging' without implementing adjustments, such as standardization, to mitigate age-related discrepancies (30).

Our results indicate that a one-unit increase in embryo score, independent of other clinical variables analyzed, led to improved clinical outcomes. The most significant effect in treatments with patient oocytes had been previously reported for other machine learning-based models using morphokinetic parameters (31). Our study also demonstrated that the AI model performed better in older patients, likely because euploid embryos tended to receive higher scores than aneuploid ones. Similarly, other AI-based models had shown that chromosomally normal embryos receive higher scores (32–34). This means that in treatments where chromosomal abnormalities are more likely, the algorithm's ability to differentiate between embryos is especially beneficial.

One of the criticisms of AI models is that they are often perceived as 'black boxes, meaning that some of their processes are not fully interpretable (17). This has sparked debate about whether a biological rationale is needed to support the use of computer-generated algorithms for embryo selection (18). However, the iDAScore algorithm has been linked to various embryonic developmental variables (9,11,19–21), and our findings confirm that its automated score correlates strongly with the morphological quality as determined by embryologists. In fact, they were so correlated that the performance of the conventional methodology to predict clinical outcomes closely approached that of the artificial intelligence-based methodology. Contrary to this, a 2023 review of 20 AI-based algorithms concluded that AI consistently outperformed embryologists in predicting clinical outcomes during embryo selection (24). However, the authors acknowledged that these results were based on locally generated databases and had not undergone independent validation. Very few publications have evaluated AI models in laboratories that were not involved in their development, and in those cases, the AUC values after independent validation were similar to those reported in our study (0.6-0.7), which closely resemble conventional methodologies (31,35).

In our dataset, using the iDAScore for embryo selection resulted in slightly better outcomes compared to conventional morphology-based methods in all subpopulations except PGT-A treatments. One possible explanation for this exception could be the assisted hatching

performed on all embryos on day three of development, which may have influenced late-stage embryonic development (36) and, consequently, image analysis. In this scenario, embryologists were aware of the procedure's impact, while the algorithm, which was not trained on embryos that had undergone this procedure, did not account for it in the evaluation.

Although not all AI-based embryo selection models show the same level of agreement with embryologists (37), we observed a good degree of concordance. Agreement was higher in conventional treatments, likely because these cases involved smaller embryo cohorts with fewer high-quality blastocysts compared to donated oocyte cycles. Notably, when there was agreement between the embryologist and the AI model, the implantation rates were higher. However, due to the retrospective nature of the study, we were unable to quantify the potential improvement had the embryo selected by the AI been transferred. While relative risk is typically used in prospective studies, we calculated it here as a measure of implantation incidence, showing that the odds ratio of implantation was greater when there was agreement between the embryologist and the AI.

The main advantage of the iDAScore system, aside from its automation and objectivity, lies in its ability to assign scores across a broader range of categories. While the conventional ASEBIR methodology only uses four categories (A, B, C, and D), with only three considered suitable for transfer, the iDAScore system provides more detailed gradations (from 1 to 9.9). This allows for a more nuanced distinction between embryos with similar morphology. Our exploratory results supported this fact. On one hand, although this was not statistically significant, grade A embryos with scores below 7 presumably implanted less than grade B embryos with scores above 7 (50.4% vs. 57.3%). On the other hand, and this was statistically significant, embryos with higher scores achieved better clinical outcomes within each morphological category. However, automatic scoring may not be as helpful when there are significant morphological differences between embryos.

A major limitation of our study is its retrospective nature. To establish the true clinical impact of the iDAScore ^{TM} v2 model, a prospective randomized study is needed. Despite being a thorough external validation in an unselected ICSI population, the model's broader applicability should be approached cautiously due to the single-center design. The application of a prediction model on a wider scale should consider specific culture conditions. Furthermore, the comparison between automated and conventional grading systems could introduce bias, as the iDAScore produces a continuous outcome while ASEBIR grading is categorical. Additionally, implantation and live birth rates are influenced by various factors beyond the variables considered in this study's GEE

models. Further research is required to explore these limitations and assess the impact of direct AI-based selection in clinical practice. Finally, we emphasize that the automation of embryo selection can achieve performance at least equal to that of embryologists. As such, future prospective studies may focus on testing non-inferiority rather than superiority hypotheses.

CONCLUSIONS

In conclusion, the iDAScore ™ v2 model proved clinically efficient in automatically ranking embryos for implantation and live birth potential. Additionally, it provided higher scores for embryos of good morphological quality, euploid embryos, and those not associated with clinical miscarriage. Although its performance was similar to that of conventional methods, the automatic scoring system offered valuable information within each morphological category, enhancing consistency in embryo selection. Furthermore, the collaboration between human expertise and AI may contribute to more favorable results in assisted reproductive treatments. Future studies should aim to evaluate AI models in a prospective manner.

CRedit authorship contribution statement

Lorena Bori: Writing-review & editing, Writing-original draft, Methodology, Data curation, Formal analysis, Conceptualization.

Marco Toschi: Writing-review & editing, Supervision, Conceptualization.

Rebeca Esteve: Data curation, Investigation.

Arantza Delgado: Data curation, Investigation.

Antonio Pellicer: Resources, Project administration

Marcos Meseguer: Resources, Project administration, Supervision, Writing-review & editing.

Acknowledgements

We acknowledge the contribution of the IVF laboratory staff at IVIRMA Global Valencia for their clinical support.

References

1. Zaninovic N, Rosenwaks Z. Artificial intelligence in human invitro fertilization and embryology. Fertil Steril 2020; 114:914–20.

2. Glatstein I, Chavez-Badiola A, Curchoe CL. New frontiers in embryo selection. J Assist Reprod Genet. 2023; 40:223–34.

3. Ueno S, Berntsen J, Ito M, Uchiyama K, Okimura T, Yabuuchi A, et al. Pregnancy prediction performance of an annotation-free embryo scoring system on the basis of deep learning after single vitrified-warmed blastocyst transfer: a single-center large cohort retrospective study. Fertil Steril 2021; 116:1172–80.

4. Zhan Q, Ph D, Sierra ET, Ph D, Malmsten J, Ph D, et al. Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. Fertility and Sterility Reports 2020; 1:133–41.

5. VerMilyea M, Hall JMM, Diakiw S, Johnston A, Nguyen T, Dakka MA, et al. Camera-agnostic self-annotating Artificial Intelligence (AI) system for blastocyst evaluation. Human Reproduction Abstracts of the 36th Virtual Annual Meeting of the ESHRE 2020; 35.

6. Bormann CL, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwala R, Kandula H, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. eLife 2020; 9:1–14.

7. Miyagi Y. Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. Reprod Med Biol 2019; 18:190–203.

8. Loewke K, Cho JH, Brumar CD, Maeder-York P, Barash O, Malmsten JE, et al. Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. Fertil Steril 2022; 117:528–35.

9. Berntsen J, Rimestad J, Lassen JT, Tran D, Kragh MF. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. PLoS One 2022; 17:1–18.

10. Ezoe K, Shimazaki K, Miki T, Takahashi T, Tanimura Y, Amagai A, et al. Association between a deep learning-based scoring system with morphokinetics and morphological alterations in human embryos. Reprod Biomed Online 2022; 45:1124–32.

11. Kato K, Ueno S, Berntsen J, Kragh MF, Okimura T, Kuroda T. Does embryo categorization by existing artificial intelligence, morphokinetic or morphological embryo selection models correlate with blastocyst euploidy rates? Reprod Biomed Online 2023; 46:274–81.

12. Theilgaard Lassen J, Fly Kragh M, Rimestad J, Nygård Johansen M, Berntsen J. Development and validation of deep learning based embryo selection across multiple days of transfer. Sci Rep 2023; 13:4235.

13. Fréour T, Fleuter N, Lammers J, Splingart C, Reignier A, Barriere P. External validation of a time-lapse prediction model. Fertil Steril 2015; 103:917–22.

14. Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. 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. Fertil Steril 2017; 107:613–21.

15. Cerrillo M, Herrero L, Guillén A, Mayoral M, García-Velasco JA. Impact of Endometrial Preparation Protocols for Frozen Embryo Transfer on Live Birth Rates. Rambam Maimonides Med J 2017; 8:e0020.

16. Cozzolino M, Franasiak J, Andrisani A, Ambrosini G, Vitagliano A. "Delayed start" gonadotropin-releasing hormone antagonist protocol in Bologna poor-responders: A systematic review and meta-analysis of randomized controlled trials. European Journal of Obstetrics & Gynecology and Reproductive Biology 2020; 244:154–62.

17. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Web Site and R Package for Computing E-values. Epidemiology 2018; 29:e45–7.

18. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med 2017; 167:268–74.

19. Altman D. Practical statistics for medical research. Chapman & Hall; 1991.

20. Afnan MAM, Liu Y, Conitzer V, Rudin C, Mishra A, Savulescu J, et al. Interpretable, not black-box, artificial intelligence should be used for embryo selection. Hum Reprod Open 2021; 2021:hoab040.

21. ESHRE Working group on time-lapse technology 2020, Apter S, Ebner T, Freour T, Guns Y, Kovacic B, et al. Good practice recommendations for the use of time-lapse technology. Hum Reprod Open 2020; 4:hoz025.

22. Diakiw SM, Hall JMM, VerMilyea M, Lim AYX, Quangkananurug W, Chanchamroen S, et al. An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos. Reprod Biomed Online 2022; 45:1105–17.

23. Joo K, Nemes A, Dudas B, Berkes-Bara E, Murber A, Urbancsek J, et al. The importance of cytoplasmic strings during early human embryonic development. Front Cell Dev Biol 2023; 11.

24. Salih M, Austin C, Warty RR, Tiktin C, Rolnik DL, Momeni M, et al. Embryo selection through artificial intelligence versus embryologists: a systematic review. Hum Reprod Open 2023; 2023:hoad031.

25. Ueno S, Berntsen J, Ito M, Okimura T, Kato K. Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: a single-centre, large-cohort retrospective study. J Assist Reprod Genet 2022; 39:2089–99.

26. Sarandi S, Boumerdassi Y, O'Neill L, Puy V, Sifer C. Interest of iDAScore (intelligent Data Analysis Score) for embryo selection in routine IVF laboratory practice: Results of a preliminary study. Gynecologie Obstetrique Fertilite et Senologie 2023; 51:372–7.

27. Zhu J, Wu L, Liu J, Liang Y, Zou J, Hao X, et al. External validation of a model for selecting day 3 embryos for transfer based upon deep learning and time-lapse imaging. Reprod Biomed Online 2023; 47:103242.

28. Ahlström A, Berntsen J, Johansen M, Bergh C, Cimadomo D, Hardarson T, et al. Correlations between a deep learning-based algorithm for embryo evaluation with

cleavage-stage cell numbers and fragmentation. Reprod Biomed Online 2023; 47:103408.

29. Ueno S, Berntsen J, Okimura T, Kato K. Improved pregnancy prediction performance in an updated deep-learning embryo selection model: a retrospective independent validation study. Reprod Biomed Online 2023; 103308.

30. Johansen MN, Parner ET, Mikkel ·, Kragh F, Kato K, Ueno S, et al. Comparing performance between clinics of an embryo evaluation algorithm based on time-lapse images and machine learning. J Assist Reprod Genet 2023; 40:2129–37.

31. Bori L, Meseguer F, Valera MA, Galan A, Remohi J, Meseguer M. The higher the score, the better the clinical outcome: Retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories. Human Reproduction 2022; 37:1148–60.

32. Barnes BS BJ, Brendel MEng M, Rajendran SB, Kim MEng J, Zisimopoulos P, Sigaras A, et al. Department of Physiology and A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: a retrospective model development and validation study. 2023.

33. Cimadomo D, Chiappetta V, Innocenti F, Saturno G, Taggi M, Marconetto A, et al. Towards Automation in IVF: Pre-Clinical Validation of a Deep Learning-Based Embryo Grading System during PGT-A Cycles. J Clin Med 2023; 12:1806.

34. Diakiw SM, Hall JMM, VerMilyea MD, Amin J, Aizpurua J, Giardini L, et al. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. Human Reproduction 2022; 37:1746–59.

35. Reignier A, Girard J, Lammers J, Chtourou S, Lefebvre T. Performance of Day 5 KIDScore ™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer. J Assist Reprod Genet 2019; 36:2279–85.

36. Cimadomo D, Capalbo A, Levi-Setti PE, Soscia D, Orlando G, Albani E, et al. Associations of blastocyst features, trophectoderm biopsy and other laboratory practice with post-warming behavior and implantation. Human Reproduction 2018; 33:1992–2001.

37. Zaninovic N, Sierra JT, Malmsten JE, Rosenwaks Z. Embryo ranking agreement between embryologists and artificial intelligence algorithms. F and S Science 2023; 5:50–7.

FIGURE CAPTIONS

Figure 1. Implantation rate and live birth rate according to automatic embryo score by quartiles.

TABLES

Table 1. Demographic table of treatments divided by subtypes. PGT-A: preimplantation genetic testing for aneuploidy; ICSI: intracytoplasmic sperm injection; BMI: body mass index; MII: metaphase II; pn: pronucleus.

	Non PGT-A (n=6,015)		PGT-A (n=1,337)	Total cycles (n=7,352)
	Conventional ICSI cycles (n=2,700)	Oocyte donation program (n=3,315)		
Oocyte Age*	36.9±4.4	25.3±4.7	35.9±5.7	31.5±7.4
BMI of patients*	23.2±3.9	23.3±3.9	23.2±3.6	23.2±3.8
Number of MII oocytes*	7.7±5.3	11.3±2.9	11.2±6.0	9.9±4.8
Number of fertilized oocytes (2pn) *	5.5±4.3	8.7±2.7	8.5±4.8	7.5±4.1
Total number of scored embryos	19,222	35,357	13,892	68,471
Single Embryo Transfer**	1,832	4,656	1,690	8,178
Double Embryo Transfer**	482	1,264	130	1,876
Transfer on D5**	1,988	5,226	1,459	8,673
Transfer on D6**	326	694	361	1,381

Fresh Embryo Transfer**	957	2,597	-	3,554
Frozen Embryo Transfer**	1,357	3,323	1,820	6,500

Table 2. Generalized estimating equations (GEE) evaluating the correlation between iDAScore and implantation as well as live birth outcomes, controlling for other potential confounders. The table displays only those variables with statistical significance ($p<0.05$) of variable-outcome association.

	Implantation				Live Birth			
	OR[a]	95% CI	p-value	e-value[b]	OR[a]	95% CI	p-value	e-value[b]
Patients (n=2,314)								
Automatic embryo score	1.31	[1.25-1.36]	<0.001	1.55	1.27	[1.21-1.33]	<0.001	1.51
Oocyte age	0.94	[0.91-0.96]	<0.001	1.21	0.93	[0.90-0.96]	<0.001	1.23
Oocyte Donation Program (n=5,920)								
Automatic embryo score	1.17	[1.14-1.20]	<0.001	1.38	1.16	[1.13-1.19]	<0.001	1.37
EmbryoTransfer Fresh (43.9%) vs. Frozen (56.1%)	0.63	[0.56-0.71]	<0.001	1.83	0.66	[0.58-0.74]	<0.001	1.76
Culture strategy Individual (62%) vs. Group (38%)	0.82	[0.72-0.93]	0.005	1.44	0.83	[0.73-0.94]	0.003	1.43
PGT-A (n=1,820)								
Automatic embryo score	-	-	-		1.05	[1-1.10]	0.044	1.18
Day of transfer D5 (80%) vs. D6 (20%)	1.35	[1.02-1.79]	0.037	1.6	-	-	-	
BMI	-	-	-		0.97	[0.94-1]	0.024	1.14

[a]ORs computed from GEE.

[b]Minimum strength of association, on the risk-ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome to fully explain away the variable–outcome association.

- variables not considered by the model.

Table 3. Performance metrics for the generalized estimating equations (GEEs) of the two classification systems, iDAScore and morphological evaluation, for implantation (a) and live-birth (b) prediction. AUC: Area under the curve. OR: odds ratio.

a) IMPLANTATION					
Treatment type		ASEBIR		iDAScore v2	
		OR [95% CI]	AUC (95% CI)	OR [95% CI]	AUC (95% CI)
PATIENT OOCYTES	<35 (n=1,117)	3.7 [2.4-5.8] (A vs. C) 2.1 [1.4-3.0] (B vs. C)	0.654 (0.623-0.686)	1.3 [1.2-1.4]	0.692 (0.661-0.722)
	35-37 (n=877)	5.5 [3.2-9.5] (A vs. C) 2.3 [1.5-3.4] (B vs. C)	0.657 (0.621-0.693)	1.3 [1.2-1.4]	0.673 (0.637-0.708)
	38-40 (n=236)	6.5 [2.3-18.1] (A vs. C) 2.4 [1.1-5.2] (B vs. C)	0.697 (0.628-0.767)	1.3 [1.1-1.5]	0.714 (0.646-0.783)
	>40 (n=84)	NS	-	1.5 [1.1-2.1]	0.772 (0.654-0.890)
OOCYTE DONATION (n=5,920)		2.6 [2.2-3.2] (A vs. C) 1.7 [1.4-2.0] (B vs. C)	0.641 (0.627-0.655)	1.2 [1.1-1.2]	0.646 (0.632-0.660)
PGT-A (n=1,820)		2.0 [1.2-3.4] (A vs. C) 1.7 [1.4-2.1] (B vs. C)	0.574 (0.547-0.600)	NS	-
b) LIVE BIRTH					
Treatment type		ASEBIR		iDAScore v2	
		OR [95% CI]	AUC (95% CI)	OR [95% CI]	AUC (95% CI)
PATIENT OOCYTES	<35 (n=1,117)	3.2 [2.1-5.0] (A vs. C) 1.7 [1.2-2.6] (B vs. C)	0.657 (0.625-0.689)	1.3 [1.2-1.3]	0.677 (0.646-0.708)
	35-37 (n=877)	5.3 [3.1-9.1] (A vs. C) 2.2 [1.4-3.4] (B vs. C)	0.652 (0.615-0.689)	1.2 [1.2-1.3]	0.654 (0.617-0.691)
	38-40 (n=236)	6.6 [2.1-21.1] (A vs. C) 2.4 [1.0-5.7] (B vs. C)	0.707 (0.630-0.784)	1.4 [1.2-1.6]	0.722 (0.650-0.794)
	>40 (n=84)	35.7 [2.3-546.7] (A vs. C) 8.2 [1.4-48.7] (B vs. C)	0.817 (0.698-0.936)	2.1 [1.5-3.1]	0.894 (0.824-0.963)
OOCYTE DONATION (n=5,920)		2.4 [2.0-3.0] (A vs. C) 1.7 [1.4-2.0] (B vs. C)	0.629 (0.615-0.643)	1.2 [1.1-1.2]	0.636 (0.622-0.650)
PGT-A (n=1,820)		1.9 [1.1-3.3] (A vs. C) 1.8 [1.5-2.3] (B vs. C)	0.586 (0.560-0.612)	1.1 [1.0-1.1]	0.558 (0.531-0.584)

SUPPLEMENTARY DATA

Supplemental Figure 1. Distribution of automatically scored embryos according to two outcomes. (a) implantation success, (b) live birth.

Supplemental Figure 2. Distribution of automatically scored embryos according to conventional morphology. ASEBIR: morphological criteria according to the association for the study of reproductive biology (A- high quality to D-low quality and E-excluded embryos).

Supplemental Figure 3. Distribution of automatically scored embryos according to ploidy status.

Supplemental Figure 4. Euploidy rate according to automatic embryo score by quartiles. Different letters indicate statistical significance between groups ($p < 0.05$).

Supplemental Figure 5. Distribution of automatically scored embryos according to miscarriage.

Supplemental Figure 6. Miscarriage rate according to automatic embryo score by quartiles. Different letters indicate statistical significance between groups ($p < 0.05$).

Supplemental Figure 7. Implantation rate according to morphological subcategories based on iDAScore. A, B, and C are the morphological categories according to ASEBIR. The symbol "+" means an automatic score > 7 and the symbol "-" means an automatic score ≤ 7 from iDAScore v2. Different letters indicate statistical significance between groups ($p < 0.05$).

Table 1.

	Non PGT-A (n=6,015)		PGT-A (n=1,337)	Total cycles (n=7,352)
	Conventional ICSI cycles (n=2,700)	Oocyte donation program (n=3,315)		
Oocyte Age*	36.9±4.4	25.3±4.7	35.9±5.7	31.5±7.4
BMI of patients*	23.2±3.9	23.3±3.9	23.2±3.6	23.2±3.8
Number of MII oocytes*	7.7±5.3	11.3±2.9	11.2±6.0	9.9±4.8
Number of fertilized oocytes (2pn) *	5.5±4.3	8.7±2.7	8.5±4.8	7.5±4.1
Total number of scored embryos	19,222	35,357	13,892	68,471
Single Embryo Transfer**	1,832	4,656	1,690	8,178
Double Embryo Transfer**	482	1,264	130	1,876
Transfer on D5**	1,988	5,226	1,459	8,673
Transfer on D6**	326	694	361	1,381
Fresh Embryo Transfer**	957	2,597	-	3,554
Frozen Embryo Transfer**	1,357	3,323	1,820	6,500

*Mean and SD represented by cycles.
**Number of KID embryos.

Table 2. Generalized estimating equations (GEE) evaluating the correlation between iDAScore and implantation as well as live birth outcomes, controlling for other potential confounders. The table displays only those variables with statistical significance (p<0.05) of variable-outcome association.

	Implantation				Live Birth			
	OR[a]	95% CI	p-value	e-value[b]	OR[a]	95% CI	p-value	e-value[b]
Patients (n=2,314)								
Automatic embryo score	1.31	[1.25-1.36]	<0.001	1.55	1.27	[1.21-1.33]	<0.001	1.51
Oocyte age	0.94	[0.91-0.96]	<0.001	1.21	0.93	[0.90-0.96]	<0.001	1.23
Oocyte Donation Program (n=5,920)								
Automatic embryo score	1.17	[1.14-1.20]	<0.001	1.38	1.16	[1.13-1.19]	<0.001	1.37
EmbryoTransfer Fresh (43.9%) vs. Frozen (56.1%)	0.63	[0.56-0.71]	<0.001	1.83	0.66	[0.58-0.74]	<0.001	1.76
Culture strategy Individual (62%) vs. Group (38%)	0.82	[0.72-0.93]	0.005	1.44	0.83	[0.73-0.94]	0.003	1.43
PGT-A (n=1,820)								
Automatic embryo score	-	-	-		1.05	[1-1.10]	0.044	1.18
Day of transfer D5 (80%) vs. D6 (20%)	1.35	[1.02-1.79]	0.037	1.6	-	-	-	
BMI	-	-	-		0.97	[0.94-1]	0.024	1.14

[a]ORs computed from GEE.

[-] variables not considered by the model.

[b]Minimum strength of association, on the risk-ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome to fully explain away the variable– outcome association.

Table 3. Performance metrics for the generalized estimating equations (GEEs) of the two classification systems, iDAScore and morphological evaluation, for implantation (a) and live-birth (b) prediction. AUC: Area under the curve. OR: odds ratio. CI: Confidence interval

a) IMPLANTATION					
Treatment type		ASEBIR		iDAScore v2	
		OR [95% CI]	AUC (95% CI)	OR [95% CI]	AUC (95% CI)
PATIENT OOCYTES	<35 (n=1,117)	3.7 [2.4-5.8] (A vs. C) 2.1 [1.4-3.0] (B vs. C)	0.654 (0.623-0.686)	1.3 [1.2-1.4]	0.692 (0.661-0.722)
	35-37 (n=877)	5.5 [3.2-9.5] (A vs. C) 2.3 [1.5-3.4] (B vs. C)	0.657 (0.621-0.693)	1.3 [1.2-1.4]	0.673 (0.637-0.708)
	38-40 (n=236)	6.5 [2.3-18.1] (A vs. C) 2.4 [1.1-5.2] (B vs. C)	0.697 (0.628-0.767)	1.3 [1.1-1.5]	0.714 (0.646-0.783)
	>40 (n=84)	NS	-	1.5 [1.1-2.1]	0.772 (0.654-0.890)
OOCYTE DONATION (n=5,920)		2.6 [2.2-3.2] (A vs. C) 1.7 [1.4-2.0] (B vs. C)	0.641 (0.627-0.655)	1.2 [1.1-1.2]	0.646 (0.632-0.660)
PGT-A (n=1,820)		2.0 [1.2-3.4] (A vs. C) 1.7 [1.4-2.1] (B vs. C)	0.574 (0.547-0.600)	NS	-
b) LIVE BIRTH					
Treatment type		ASEBIR		iDAScore v2	
		OR [95% CI]	AUC (95% CI)	OR [95% CI]	AUC (95% CI)
PATIENT OOCYTES	<35 (n=1,117)	3.2 [2.1-5.0] (A vs. C) 1.7 [1.2-2.6] (B vs. C)	0.657 (0.625-0.689)	1.3 [1.2-1.3]	0.677 (0.646-0.708)
	35-37 (n=877)	5.3 [3.1-9.1] (A vs. C) 2.2 [1.4-3.4] (B vs. C)	0.652 (0.615-0.689)	1.2 [1.2-1.3]	0.654 (0.617-0.691)
	38-40 (n=236)	6.6 [2.1-21.1] (A vs. C) 2.4 [1.0-5.7] (B vs. C)	0.707 (0.630-0.784)	1.4 [1.2-1.6]	0.722 (0.650-0.794)
	>40 (n=84)	35.7 [2.3-546.7] (A vs. C) 8.2 [1.4-48.7] (B vs. C)	0.817 (0.698-0.936)	2.1 [1.5-3.1]	0.894 (0.824-0.963)
OOCYTE DONATION (n=5,920)		2.4 [2.0-3.0] (A vs. C) 1.7 [1.4-2.0] (B vs. C)	0.629 (0.615-0.643)	1.2 [1.1-1.2]	0.636 (0.622-0.650)
PGT-A (n=1,820)		1.9 [1.1-3.3] (A vs. C) 1.8 [1.5-2.3] (B vs. C)	0.586 (0.560-0.612)	1.1 [1.0-1.1]	0.558 (0.531-0.584)