

# ASSESSING SEX-BASED DISPARITIES IN EMBRYO SELECTION ALGORITHMS: A COMPREHENSIVE ANALYSIS OF MORPHOKINETIC PARAMETERS AND ALGORITHMIC FAIRNESS

Teodora Popa, Chloe He, Matthew Lau, Christian Ottolini, Helen O'Neill

## Background

Artificial intelligence algorithms are increasingly utilised for embryo selection through morphokinetics assessment (e.g. tPNf, t2-t8, tB, ICM and TE grades) to predict embryo viability or likelihood of live birth. However, these models assume similar morphokinetics irrespective of embryo sex<sup>1</sup>. With conflicting evidence on disparities between XX and XY embryos' morphokinetics<sup>2-5</sup>, our study examines the relationship between embryo sex and the scores assigned by three selection algorithms: manual Gardner-grading system, traditional machine-learning, and deep-learning models. Our primary focus is whether the predictions of embryo selection algorithms themselves are affected by embryo sex, potentially impacting the embryo selection accuracy and fairness.

## Methods

In the largest study on sex differences in preimplantation embryos (n=1411), sex and ploidy data were sourced at the expanded blastocyst stage via PGT-A. Continuous culturing occurred in Geri incubators, with morphokinetic parameters derived using CHLOE EQ. Blastocysts were graded using manual grading (Gardner system<sup>6</sup>), KIDScore D37 (traditional machine learning), and CHLOE EQ8 (deep learning), capturing the main grading methods currently used. Mann-Whitney U and chi-squared tests compared XX and XY gradings, with detailed analyses on morphokinetics (tPNf, t2-t9+, cc2, cc3, tM, tSB-tEB, ICM and TE grades).

## Results

XY embryos were more likely than XX embryos to score higher under standard morphological grading ( $p < 0.00001$ ) and the KIDScore D3 ( $p = 0.0182$ ) – although differences were marginal and the overlap between the sexes substantial enough to preclude reliable sex selection. Discrepancies in standard grading were largely due to trophoctoderm grade differences, but reasons for KIDScore D3 disparities remain uncertain. No difference between XX and XY embryos was identified in CHLOE EQ scores ( $p = 0.208$ ).

## Conclusions

We studied three embryo selection algorithms to assess potential sex bias. Two showed slight biases towards XY embryos. Recognising this bias is essential for fair and accurate embryo selection processes and to mitigate any influence ART might have on population dynamics.