LETTERS

CLARIFICATION

An improvement in the method used to assess discriminatory ability when predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation

David J McLernon *research fellow in medical statistics*[1], Ewout W Steyerberg *professor of medical decision making*[2], Egbert R te Velde *emeritus professor of reproductive medicine*[2], Amanda J Lee *professor of medical statistics*[3], Siladitya Bhattacharya *professor of reproductive medicine and director*[3]

[1]Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK; [2]Department of Public Health, Erasmus MC-University Medical Centre Rotterdam, 3000 CA Rotterdam, The Netherlands; [3]Institute of Applied Health Sciences, Polwarth Building, University of Aberdeen, Aberdeen, AB25 2ZD, UK

Summary

Our paper presents two clinical prediction models that estimate the chance of having a baby over multiple complete cycles of in vitro fertilisation (IVF)—that is, cumulative live birth.[1] The pretreatment model predicts the chance of cumulative live birth before treatment starts, and the post-treatment model predicts the chance of cumulative live birth just after the first embryo transfer. Through a collaboration with researchers from the University of Utrecht, who have externally validated these models, we have decided to revise the method used to assess the discriminatory ability of our models in the original study. In time to event models, such as ours, discrimination indicates the proportion of all pairs of women who can be ordered such that the woman with the lower predicted chance of live birth is the one who either did not have a live birth or had more complete cycles of IVF to have a live birth. Discrimination is assessed using the C index, where 1 is perfect discrimination and 0.5 is no better than a coin toss.[2]

We request this amendment so that other researchers externally validating these models can use this revised approach to calculate the C index. They can then compare the discriminatory ability of their own cohort with the model applied to the original development cohort (as revised and presented here).

Revised method

The model used in our analysis was a discrete time logistic regression model. This is a type of time to event model used when the time measurement is not continuous, such as IVF cycle number.[1] Three elements are needed to calculate the C index for this model: the observed live birth status, the predicted probability of live birth, and the number of complete cycles to end of follow-up (the discrete time variable, which ranged from 1 to 6) for each woman. In the original study, the C index was estimated using the predicted probability of live birth at the last observed complete cycle for each woman. We noted, however, that for a Cox time to event model, which is used when time is continuous (calendar time), one can use the linear predictor from the model instead of the predicted probability.[3] The linear predictor is the equivalent of using the predicted probability at the same time point, rather than the last observed time point, for all women.

The following example shows where our original method does not discriminate correctly for a small proportion of pairs. Suppose we have a pair of women, one of whom had a live birth after her second complete cycle and the other had three complete cycles that ended without a live birth. For the first woman, the model gave predicted probabilities of 20% at complete cycle 1 and 23% up to and including complete cycle 2. For the second woman her predicted probabilities of live birth were 18%, 22%, and 24% up to and including complete cycles 1, 2, and 3, respectively. In our original method, which used probabilities at the last observed complete cycle, we would have said that our model discriminated poorly for this pair because it gave a higher probability to the second woman. But when we compare the two women at the same time point (complete cycle 1), the

d.mclernon@abdn.ac.uk

model correctly gave a higher predicted probability to the first woman, who did have a live birth.

In a Cox model each patient has one linear predictor value, whereas in a discrete time logistic regression model the linear predictor varies for each discrete time point (complete cycle number). For calculation of the C index, the linear predictor value serves to determine the order of the women in terms of their prognostic chances. So if the same complete cycle number is used to calculate the linear predictor value for all women, the ordering will remain the same no matter which complete cycle number we use. We recalculated the C index for both the pretreatment and post-treatment models using the linear predictor at complete cycle 1.

Revised result

For the pretreatment model the C index decreased from 0.73 (95% confidence interval 0.72 to 0.74) to 0.69 (0.68 to 0.69), and for the post-treatment model it increased from 0.72 (0.71 to 0.73) to 0.76 (0.75 to 0.77). These changes do not affect any of the model estimates, predictions, or overall conclusions in the original paper. They do, however, provide a more robust method and estimate of discriminatory ability, which other researchers may use in future validation studies of our models.

Competing interests: None declared.

1   McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ* 2016;355:i5735. 10.1136/bmj.i5735 27852632
2   Harrell FEJr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87. 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4 8668867
3   Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004;23:2109-23. 10.1002/sim.1802 15211606