Human Reproduction, Vol.0, No.0 pp. 1–2, 2016
doi:10.1093/humrep/dew105

human reproduction

SUPPLEMENTARY DATA

Imputation of missing data

We used single stochastic imputation based on patients without major protocol violation and with initial hCG >25 IU/l. If the blood sample for the 48-h hCG was taken more than 3 days after the blood sample to quantify initial hCG, the 48-h hCG level was considered missing and was also imputed. We used the method of fully conditional specification (van Buuren, 2007), which was implemented in SAS v9.4 using PROC MI. Variables included in the imputation analysis were outcome (PUL, pregnancy of unknown location; PPUL, persistent PUL; EP, ectopic pregnancy; FPUL, failed PUL; IUP, intrauterine pregnancy), age, vaginal bleeding, the logarithm of initial hCG, the logarithm of 48-h hCG, the logarithm of initial progesterone and centre. Some patients were lost to follow up and therefore had missing outcome. Based on recommendations from recent research (Sullivan *et al.*, 2015), the outcome of these patients was imputed, and these patients were included in the analysis of the imputed data. Therefore, the imputed dataset contained 914 patients.

This approach assumes that missing data have occurred under the 'missing at random' (MAR) mechanism, which means that missing data are missing randomly conditional on the available data (Sterne *et al.*, 2009). Missing outcome data may not be completely random, as perhaps women with a FPUL or IUP are more often lost to follow-up because the pregnancy has terminated or is evolving without any issues or symptoms. Given that progesterone values and hCG patterns are clearly different for the different PUL outcomes, we believe that conditioning imputation on this information makes missing outcome data largely MAR.

Diagnostic performance of the M4 risk model

Discrimination

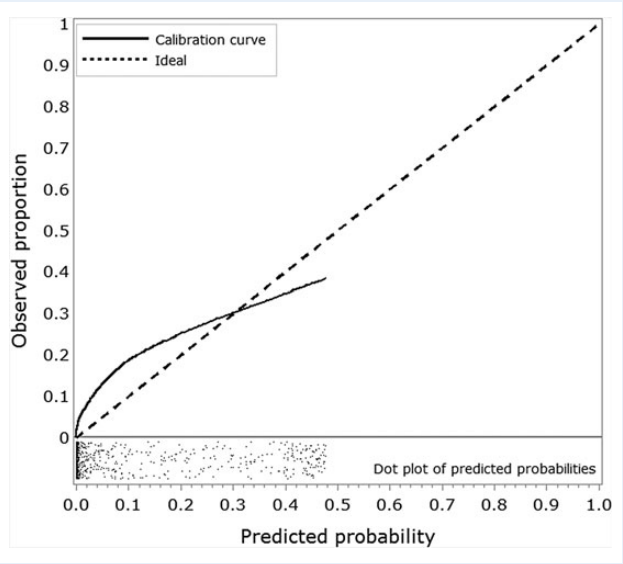
The Polytomous Discrimination Index (PDI) is an extension of the area under the receiver operating characteristic curve (AUC) for outcomes with more than two categories. The PDI for cases with complete outcome data was 0.82, and for the imputed data was 0.80 (Supplementary Table SI). This means that, given one FPUL, and IUP, and one EP, on average 2.5 (complete cases) or 2.4 (imputed data) of these three patients are correctly identified by the model (Van Calster *et al.*, 2012). PDI for individual centres varied between 0.82 and 0.84 for complete cases, and between 0.79 and 0.82 for imputed data.

The AUC for EP versus FPUL/IUP was 0.85 (95% confidence interval 0.80–0.88) for complete cases and 0.84 (0.79–0.87) for imputed data (Supplementary Table SI). The AUC for individual centres varied between 0.84 and 0.85 for complete cases, and between 0.81 and 0.85 for imputed data. The AUC for FPUL versus IUP was extremely high for all centres (0.98 overall for complete cases, 0.97 overall for imputed data) (Supplementary Table SI).

Supplementary Table SI Discrimination performance of the M4 model

Subgroup	PDI	AUC EP versus FPUL/IUP (95% CI)	AUC FPUL versus IUP (95% CI)
Complete cases ($n = 835$)			
All PUL	0.82	0.85 (0.80–0.88)	0.98 (0.96–0.98)
Queen Charlotte's and Chelsea	0.82	0.85 (0.79–0.89)	0.97 (0.95–0.98)
Chelsea and Westminster	0.84	0.85 (0.69–0.94)	0.98 (0.95–0.99)
Hillingdon	0.82	0.84 (0.76–0.90)	0.99 (0.96–0.99)
All PUL without violations	0.86	0.88 (0.84–0.92)	0.98 (0.97–0.99)
Imputed data ($n = 914$)			
All PUL	0.80	0.84 (0.79–0.87)	0.97 (0.95–0.98)
Queen Charlotte's and Chelsea	0.80	0.85 (0.79–0.89)	0.95 (0.93–0.97)
Chelsea and Westminster	0.79	0.81 (0.66–0.90)	0.97 (0.94–0.99)
Hillingdon	0.82	0.83 (0.75–0.90)	0.99 (0.96–0.99)
All PUL without violations	0.85	0.88 (0.83–0.91)	0.98 (0.96–0.99)

PUL, pregnancy of unknown location; EP, ectopic pregnancy; FPUL, failed PUL; IUP, intrauterine pregnancy; PDI, Polytomous Discrimination Index; AUC, area under the receiver operating characteristic curve; CI, confidence interval.



Supplementary Figure S1. Flexible calibration curve for the predicted probability of ectopic pregnancy based on imputed data. The calibration curve based on complete cases is nearly identical (not shown). At the bottom of the figure, a dot plot of the predicted probabilities is given. The dot plot shows that many pregnancy of unknown location (PUL) have a predicted probability that is very close to 0, and that no PUL has a predicted probability of 0.5 or higher.

Calibration

The flexible calibration curves are similar for complete cases and for imputed data (Supplementary Fig. S1). The predicted risk for EP was never higher than 0.5. For predicted risks up to 0.3, the calibration curves indicated underestimation. For example, for patients with a predicted risk of 0.05 (5%), the observed proportion of EP is somewhere between 10 and 15%. For predicted risks above 0.3, which are not very common, the analysis suggests some overestimation. The observation of underestimation of low predicted risks and overestimation of high predicted risks suggests that the M4 model suffers from overfitting of the risk of EP.

References

Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**:b2393.

Sullivan TR, Salter AB, Ryan P, Lee KJ. Bias and precision of the "multiple imputation, then deletion" method for dealing with missing outcome data. *Am J Epidemiol* 2015;**182**:528–534.

van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;**16**:219–242.

Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the Polytomous Discrimination Index. *Stat Med* 2012; **31**:2610–2626.