Appendix 1. Performance of Logistic Regression Models: Preterm Delivery Within 7 Days Predicted With Cervical Length and Fetal Fibronectin Results

The performance of the 3 prediction models including cervical length as a predictor was compared in terms of overall fit, discrimination, calibration and reclassification. Overall fit of the models was expressed with Nagelkerke R^2, Brier and *scaled* Brier score. Improvement in fit by adding fibronectintest results was tested with the X^2 test. The ability of the models to discriminate between women who delivered within 7 days and those who delivered at least 7 days after testing was expressed as the area under the receiver operating characteristics curve (AUC) and as discrimination slopes. Agreement between predicted and observed proportions of women with spontaneous preterm delivery within 7 days after enrollment, also known as calibration, was visualized in a calibration plot and miscalibration was tested for significance with the Hosmer-Lemeshow test statistic (1).

Improvement in reclassification between the model based on cervical length only and the model including fibronectin results was expressed as Net Reclassification Improvement (NRI) and integrated discrimination improvement (IDI). One needs to define a decision threshold, indicating high and low risk, before calculating the NRI. The NRI is the sum of the difference in proportion of individuals with the outcome moving up (above the 5% risk) minus the proportion of those moving down (beneath the 5% risk), and the proportion of individuals without the outcome moving down minus the proportion in those moving up. The IDI integrates the NRI over all possible cut-offs, and is the equivalent to difference in discrimination slopes (1)

Reference List

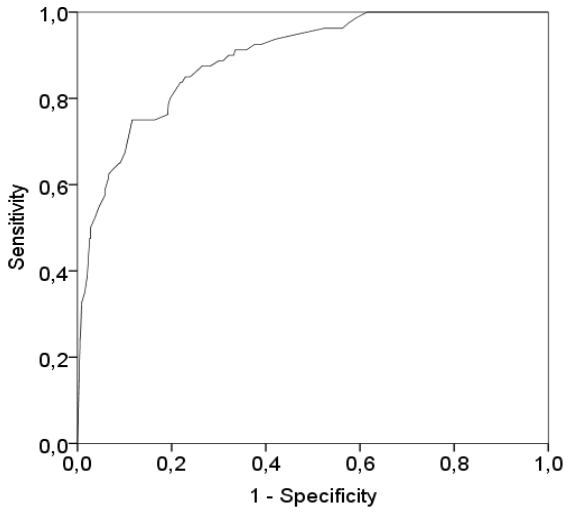
1. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-38.

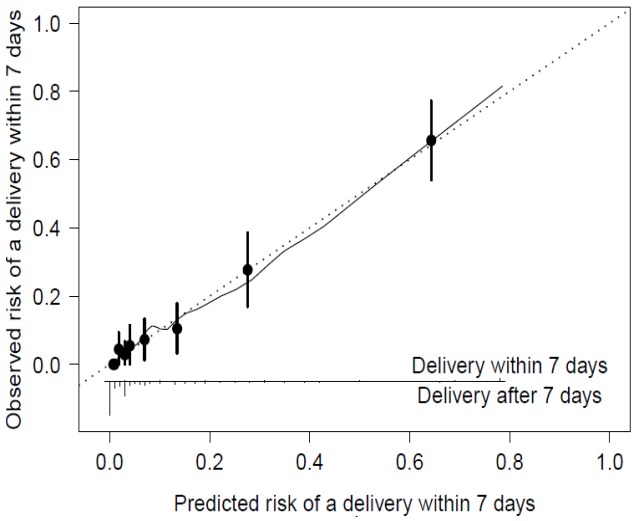
van Baaren G-J, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM et al. Predictive value of cervical length measurement and fibronectin testing in threatened preterm labor. Obstet Gynecol 2014;123.

The authors provided this information as a supplement to their article.

Performance measure	Cervical length	Fetal fibronectin and cervical length	Fetal fibronectin and cervical length with interaction
Overall fit			
Brier	0.070	0.069	0.069
Brier scaled	33.1%	34.9%	34.6%
R^2 (Nagelkerke)	41.2%	45.2%	44.9%
Discrimination			
C-stat (95%CI)	0.87 (0.83 to 0.91)	0.89 (0.86 to 0.93)	0.89 (0.86 to 0.93)
Discrimination slope	0.3331	0.3621	0.3483
Calibration			
Calibration in the large	0.0043	0.0088	-0.0281
Calibration slope	0.9918	0.9977	0.9717
Hosmer-Lemeshow test	X^2= 10.668 p= 0.055	X^2= 6.653 p=0.354	X^2=12.143 p= 0.059
Reclassification			
IDI	x	1.8%	1.5%
NRI	x	0.099	0.091

Receiver operating curve and calibration plot of the model including cervical length and fetal fibronectin results (without interaction term); the model with the best performances as been shown in the table above.





van Baaren G-J, Vis JY, Wilms FF, Oudijk MA, Kwee A, Porath MM et al. Predictive value of cervical length measurement and fibronectin testing in threatened preterm labor. Obstet Gynecol 2014;123.

The authors provided this information as a supplement to their article.

