*Ultrasound Obstet Gynecol* 2017; **49**: 169–170
Published online in Wiley Online Library (wileyonlinelibrary.com).

# Referee Commentaries

**Re: Validation of a first-trimester screening model for pre-eclampsia in an unselected population.** E. Scazzocchio, F. Crovetto, S. Triunfo, E. Gratacós and F. Figueras. *Ultrasound Obstet Gynecol* 2017; **49**: 188–193.

The authors of this article have published previously a first-trimester predictive model for pre-eclampsia based on maternal history and body mass index, uterine artery Doppler, blood pressure and pregnancy-associated plasma protein-A[1]. In the current study they performed a temporal validation of the model, i.e. they assessed the model's performance in their own institution in more recent years. They found a predictive performance similar to that of the construction cohort.

Prediction models may perform differently in populations that are not the same as the one used originally in their construction. Moreover, it is not only the discrimination power of the model (e.g. sensitivity, specificity, receiver–operating characteristics curves) that must be reported, but also its calibration, i.e. the agreement between predicted and observed outcomes. A systematic review[2] found that only 20% of predictive models for pre-eclampsia reported in the literature had internal validation, and even fewer (7%) had external validation; calibration was assessed for only 12% of them. These figures were similarly low for other obstetric conditions. Moreover, when external validation is performed, the predictive performance can be much lower than that in the original publication, as shown by Oliveira *et al.*[3] for the first-trimester prediction of pre-eclampsia.

The current article has a robust design and shows stable performance of the authors' prediction model in the setting in which it was constructed. However, it also shows some of the limitations of temporal validation: in a few years, the population attending the same clinic had changed i.e. the ethnic mix had drifted and more women with chronic medical conditions or previous pregnancy complications were presenting for first-trimester assessment. This might help to explain the minor differences in discrimination and calibration for early pre-eclampsia between the construction and the validation cohorts, as well as the relatively poor calibration for late pre-eclampsia; late pre-eclampsia is relatively common and is known to be affected by a number of maternal factors, the distribution of which had changed over time in this setting.

This study is an example of how validation is an integral part of predictive model research. Ideally, academic journals should strive for validation to be part of the original reporting of predictive models, as it is a necessary step towards clinical implementation. Further research on the validation of existing models is also required.

F. Prefumo
*Department of Obstetrics and Gynaecology, University of Brescia, Piazzale Spedali Civili 1, 25123, Brescia, Italy*
*(e-mail: federico.prefumo@gmail.com)*
**DOI:** 10.1002/uog.17395

## References

1. Scazzocchio E, Figueras F, Crispi F, Meler E, Masoller N, Mula R, Gratacos E. Performance of a first-trimester screening of preeclampsia in a routine care low-risk setting. *Am J Obstet Gynecol* 2013; **208**: 203.e1–e10.
2. Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, Mol BW, Pajkrt E, Moons KG, Schuit E. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol* 2016; **214**: 79–90.e36.
3. Oliveira N, Magder LS, Blitzer MG, Baschat AA. First-trimester prediction of pre-eclampsia: external validity of algorithms in a prospectively enrolled cohort. *Ultrasound Obstet Gynecol* 2014; **44**: 279–285.