Externally validated prediction models for pre-eclampsia

Tiruneh, S. A.; Vu, T. T. T.; Moran, L. J.; Callander, E. J.; Allotey, J.; Thangaratinam, S.; Rolnik, D. L.; Teede, H. J.; Wang, R.; Enticott, J.

DOI:
10.1002/uog.27490

Ultrasound Obstet Gynecol 2024; **63**: 592–604
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.27490.

Externally validated prediction models for pre-eclampsia: systematic review and meta-analysis

S. A. TIRUNEH[1], T. T. T. VU[1], L. J. MORAN[1], E. J. CALLANDER[1,2], J. ALLOTEY[3,4], S. THANGARATINAM[3,4,5], D. L. ROLNIK[6], H. J. TEEDE[1], R. WANG[6] and J. ENTICOTT[1]

[1]*Monash Centre for Health Research and Implementation, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, Australia;* [2]*School of Public Health, Faculty of Health, University of Technology Sydney, Sydney, NSW, Australia;* [3]*World Health Organization (WHO) Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK;* [4]*NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK;* [5]*Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK;* [6]*Department of Obstetrics and Gynaecology, Monash University, Clayton, VIC, Australia*

KEYWORDS: eclampsia; external validation; prediction; pre-eclampsia; prognostic

CONTRIBUTION

What are the novel findings of this work?
This systematic review revealed a lack of external validation of published prediction models for pre-eclampsia (PE), apart from the triple-test Fetal Medicine Foundation (FMF) model that had been validated externally in numerous high- and middle-income settings. The FMF model demonstrated outstanding discrimination and calibration in preterm PE prediction. Apart from the FMF model, the majority of existing models had poor-to-good discrimination and poor calibration performance on external validation.

What are the clinical implications of this work?
The triple-test FMF model is effective in predicting preterm PE. Future work may need to assess its feasibility and cost-effectiveness, as it requires some tests that may be cost-prohibitive in some settings. It would be ideal for low-income settings to have a comprehensive PE screening tool that is widely applicable and cost-effective, utilizing standardized prognostic factors and harmonized data sources.

ABSTRACT

Objective *This systematic review and meta-analysis aimed to evaluate the performance of existing externally validated prediction models for pre-eclampsia (PE) (specifically, any-onset, early-onset, late-onset and preterm PE).*

Methods *A systematic search was conducted in five databases (MEDLINE, EMBASE, Emcare, CINAHL and Maternity & Infant Care Database) and using Google Scholar/reference search to identify studies based on the Population, Index prediction model, Comparator, Outcome, Timing and Setting (PICOTS) approach until 20 May 2023. We extracted data using the CHARMS checklist and appraised the risk of bias using the PROBAST tool. A meta-analysis of discrimination and calibration performance was conducted when appropriate.*

Results *Twenty-three studies reported 52 externally validated prediction models for PE (one preterm, 20 any-onset, 17 early-onset and 14 late-onset PE models). No model had the same set of predictors. Fifteen any-onset PE models were validated externally once, two were validated twice and three were validated three times, while the Fetal Medicine Foundation (FMF) competing-risks model for preterm PE prediction was validated widely in 16 different settings. The most common predictors were maternal characteristics (prepregnancy body mass index, prior PE, family history of PE, chronic medical conditions and ethnicity) and biomarkers (uterine artery pulsatility index and pregnancy-associated plasma protein-A). The FMF model for preterm PE (triple test plus maternal factors) had the best performance, with a pooled area under the receiver-operating-characteristics curve (AUC) of 0.90 (95% prediction interval (PI), 0.76–0.96), and was well calibrated. The other models generally had poor-to-good discrimination performance (median AUC, 0.66 (range, 0.53–0.77)) and were overfitted on external validation. Apart from the FMF model, only two models that were validated multiple times for any-onset PE prediction, which were based on maternal characteristics*

Correspondence to: Assoc. Prof. J. Enticott, 43–51 Kanooka Grove, Clayton, VIC 3168, Australia (e-mail: joanne.enticott@monash.edu)

Accepted: 8 September 2023

only, produced reasonable pooled AUCs of 0.71 (95% PI, 0.66–0.76) and 0.73 (95% PI, 0.55–0.86).

Conclusions *Existing externally validated prediction models for any-, early- and late-onset PE have limited discrimination and calibration performance, and include inconsistent input variables. The triple-test FMF model had outstanding discrimination performance in predicting preterm PE in numerous settings, but the inclusion of specialized biomarkers may limit feasibility and implementation outside of high-resource settings.*

INTRODUCTION

Globally, pre-eclampsia (PE) affects 2–8% of pregnancies[1], with reported rates of 3% in the Americas, 5.3% in the European region, 1% in the Eastern Mediterranean region, 5.6% in the African region[2] and 3.3% in Australia[3,4]. PE causes more than 500 000 fetal and neonatal deaths, and 70 000 maternal deaths every year worldwide[5]. It is also associated with adverse neonatal[6] and maternal[7,8] outcome. Thus, early identification is required for close antenatal monitoring and prevention strategies to reduce the risk of adverse maternal and perinatal outcome. Aspirin prophylaxis for high-risk women, ideally before 16 weeks' gestation, can reduce the rates of PE by 18%, preterm birth by up to 20% and adverse birth outcome by up to 21%[9,10].

In the clinical setting, healthcare providers often attempt to predict whether a specific health event will occur in the future to their patient (prognostic setting)[11]. Risk prediction is rarely based on a single predictor and health professionals naturally integrate several signs, symptoms and biomarkers to predict a future health condition[12]. Development and validation of a primary prediction model to predict the individual risk of a future event, using a combination of multiple prognostic patient characteristics (demographic, clinical and biomarker prognostic factors), is an increasingly used approach[13]. However, the performance of a primary prediction model is often reported based on the data that were used to create the model (internal validation). Before any prediction model can be considered for clinical implementation, it requires geographic and/or temporal external validation to determine whether it will allow accurate prediction based on data from a different population that were not used to create the model[13,14].

Externally validated models that have good discrimination performance and are well calibrated for PE prediction would be of high value for clinical decision making and would help guide individualized risk stratification. This systematic review and meta-analysis aimed to evaluate the predictive performance of existing prediction models for PE that had been externally validated across different settings and populations.

METHODS

Review question and search strategy

This systematic review examined externally validated prediction models for PE. The review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42022330147). We used the Population, Index prediction model, Comparator, Outcome, Timing and Setting (PICOTS) approach for prediction studies to formulate inclusion and exclusion criteria[15]: P (pregnant women), I (externally validated index prediction models), C (not applicable), O (PE), T (prediction of PE during pregnancy) and S (individualized risk stratification for clinical set-up).

A systematic literature search was conducted in Ovid (MEDLINE, EMBASE, Emcare and Maternity & Infant Care Database (MIDIRS)) and CINAHL databases without any restriction in terms of the publication year. The last search was performed and updated on 20 May 2023. In addition, a Google Scholar gray literature search was conducted as per Enticott *et al.*[16]. Furthermore, we explored the reference list of all included studies and a recent narrative review[17]. The search strategies were developed following search filters for prognostic and diagnostic studies[18] and by consulting librarians at Monash University, Melbourne, VIC, Australia. Medical Subject Heading (MeSH) and free-text terms were used to identify potential studies (Appendix S1). Boolean operators (AND, OR and NOT) and truncation were used to combine the search strategies.

Eligibility of prediction models

Externally validated models for any-onset, early-onset, late-onset and preterm PE were included. Studies that conducted external validation of prediction models for PE using cohort/follow-up, nested case–control, case–control or randomized controlled trial design were included in this review. Prediction model studies conducted for hypertensive disorders of pregnancy and gestational hypertension alone were excluded from this review unless they also reported a separate model for PE.

Screening of prediction model search results

The Covidence platform was used to screen the search results[19]. Two reviewers (S.A.T. and T.T.T.V.) assessed independently the title and abstract of the articles after the removal of duplicates, which was followed by full-text screening. Discrepancies between the two reviewers were resolved through discussion.

Assessment of methodological quality of included studies

Risk of bias was assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST)[20] by two reviewers (S.A.T. and T.T.T.V.). The tool has four

key domains (participants, predictors, outcome and analysis) structured in 20 signaling questions to facilitate risk-of-bias assessment. Each domain was rated as having high, low or unclear risk of bias.

Data extraction

Data for included articles were extracted using the CHecklist for critical Appraisal and data extraction in systematic Reviews of clinical prediction Modelling Studies (CHARMS)[21]. Authors, publication year, source of data, predicted outcome(s), candidate prognostic factors, sample size, model development, model discrimination performance, calibration measures and results, including final multivariable models, were extracted. Two reviewers (S.A.T. and T.T.T.V.) independently extracted the data. Discrepancies were resolved through discussion between the two reviewers and with another reviewer (J.E.) if necessary.

In this review, a set of model coefficients tested using a predefined set of input variables was considered a model. The below example demonstrates this more clearly. The same set of coefficients validated externally in different sets of input data that were defined differently was counted as two models: (i) maternal data recorded at 34 weeks' gestation or earlier (early onset) and (ii) maternal data recorded after 34 weeks (late onset).

Data analysis

The descriptive synthesis of all externally validated prediction models is reported in tables and graphs. Discrimination and calibration performance of the externally validated prediction model was descriptively compared and described. Performance of the external validations of different models is visualized in forest plots so that the reader can compare the performance. An area under the receiver-operating-characteristics curve (AUC) ≤ 0.5 suggested no discrimination ability, $0.5 < \text{AUC} < 0.7$ was considered indicative of poor discrimination, $0.7 \leq \text{AUC} < 0.8$ indicated good discrimination, $0.8 \leq \text{AUC} < 0.9$ indicated excellent discrimination and AUC ≥ 0.9 was considered indicative of outstanding discrimination performance[22]. Prediction slope and prediction intercept were extracted, as both need to be reported to judge the overall calibration[23]. A model is well-calibrated[24] when the Hosmer–Lemeshow *P*-value is not significant and/or the calibration slope value approaches 1. Furthermore, the model discrimination performance was explored descriptively, according to (i) whether maternal demographic and clinical characteristics *vs* maternal demographic and clinical characteristics plus biomarkers were included and (ii) the number of prognostic factors < 5 *vs* ≥ 5 (median value) included.

Meta-analyses were performed only in studies validating the same prediction model as recommended in the current guidance[13]. We used metamisc, Meta-Analysis of Diagnosis and Prognosis Research Studies R package[25], to derive the AUC summary estimate with 95% CI and 95% prediction interval (PI). We estimated the summary discriminative performance using a random-effects model with restricted maximum likelihood method of estimation. The extracted AUC with the corresponding CI, sample size and events of each study were used to estimate the 95% CI summary estimate and PI of AUC. PIs were calculated only when the number of studies was ≥ 3, due to the required $n - 2$ degrees of freedom.

RESULTS

Externally validated prediction model selection

The details of the article selection procedure are presented in the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart (Figure 1). A total of 9381 records were retrieved from five databases. A total of 2343 duplicate records were removed, leaving 7038 records that underwent title and abstract screening. This resulted in 255 articles that were retrieved for full-text review. After screening the full text, 23 studies reporting external validation of 52 models were included. The majority of these studies (16/23) focused on external validation of the Fetal Medicine Foundation (FMF) competing-risks prediction model for preterm PE, which applies data from the first trimester (11–14 weeks) of pregnancy. The remaining 7/23 studies reported external validation of 51 models, including 20 any-onset models (Table 1), 17 early-onset models (Table S1) and 14 late-onset models (Table S2).

Characteristics of included studies and data used in external validation

The research groups that produced the seven publications that externally validated any-, early- and late-onset models[26–32] were from UK[27,30], Italy[31], France[26], USA[32] and The Netherlands[28,29]. All seven studies were at low risk of bias in terms of methodological quality (Table S3). The external validation datasets had sample sizes ranging from 1145 to 59 892 for any-onset PE. External validation datasets were from two single-center studies (France and UK), four multicenter (The Netherlands, Italy and USA) studies and one individual participant data (IPD) meta-analysis with data from multiple sources (IPD from 11 UK cohort studies with 217 415 pregnant women) (Table 1). All prediction models were validated in unselected (high or low risk for PE) pregnant women. Model regression coefficients were reported for almost all of the included models (16/20, 17/17 and 13/14 for any-, early- and late-onset PE models, respectively). The details of the primary prediction models are described in Table S4.

Characteristics of included models

Any-onset PE models

As shown in Table 1, 20 any-onset PE prediction models were validated externally[33–51] and reported in five

publications[27–31]. Three any-onset PE models[33,34,45] were validated externally three times in different cohorts[27–30], two models[42,46] were validated externally twice in different cohorts[29,30] and the rest were validated in one cohort only (Table 1). Among the original studies reporting prediction models used in the external validations, nine[33,34,37,41,43,44,47–49] were from the UK, five[39,40,42,45,46] were from the USA, two[36,38] were from Canada, two[50,51] were from multiple countries (New Zealand, Australia, UK and Ireland (the Screening for Pregnancy Endpoints (SCOPE) study)) and one[35] was from Iran. Most (90% (18/20)) of the primary any-onset PE prediction studies were conducted using a prospective-cohort/follow-up design. Among the included any-onset PE prediction models, the minimum number of prognostic factors included was two (one model) and the maximum number of prognostic factors included was 10 (one model).

Early-onset PE models

Seventeen[34,41–43,46,52–63] early-onset PE prediction models were validated externally in four publications[26,27,30,32]. The model by Scazzocchio *et al.*[59] was validated externally four times in four different cohorts[26,27,30,32], four models[46,60–62] were validated twice in different cohorts[27,30,32] and the remaining models were validated externally only once (Table S1). Among the included early-onset PE prediction models, the minimum number of prognostic factors included was two (two models) and the maximum number of prognostic factors included was 11 (one model).

Late-onset PE models

Fourteen[34,43,46,52–57,59–63] late-onset PE prediction models were validated externally and reported in four publications[26,27,30,32]. The model by Scazzocchio *et al.*[59] was validated externally four times in four different cohorts[26,27,30,32]. The models by Parra-Cordero *et al.*[60] and Poon *et al.*[61] were validated externally twice in three different cohorts[27,30,32], while the remaining models were validated externally only once (Table S2). Among the late-onset PE prediction models, the minimum number of prognostic factors included was three (four models) and the maximum number of prognostic factors included was seven (one model).

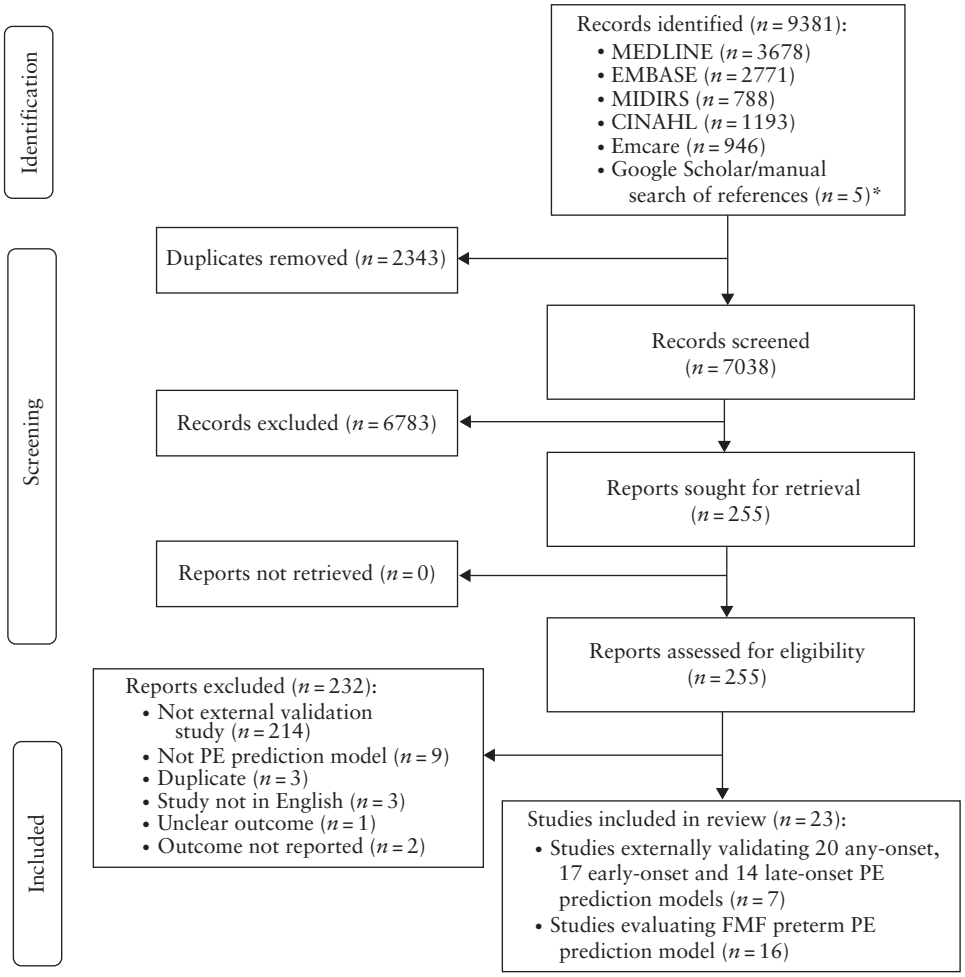

Figure 1 PRISMA flowchart summarizing inclusion of studies in systematic review. *Reference list of recent publications[72,78] and Google Scholar citation searches for the original Fetal Medicine Foundation (FMF) model. MIDIRS, Maternity & Infant Care Database; PE, pre-eclampsia.

FMF competing-risks model for preterm PE

Table 2 summarizes the characteristics of studies that validated the FMF preterm-PE prediction model. The FMF competing-risks model[64] was developed originally in 2012 and was later updated to include biochemical markers of serum pregnancy-associated plasma protein-A (PAPP-A) and placental growth factor (PlGF)[65]. The FMF competing-risks model focuses on preterm PE in the first trimester of pregnancy. The performance of the FMF competing-risks model has been evaluated in 16 studies[66–81], conducted in European countries (UK, Denmark, Belgium, Spain and The Netherlands), Asian countries (India, China and Japan), Brazil and Australia. In the preterm-PE group, the median maternal age ranged from 28 to 40 years. The sample sizes for the included studies ranged from 362 to 61 174. The incidence of term PE was generally higher than that of preterm PE, except for in studies conducted by Lobo *et al.*[81] and Hu *et al.*[76] (Table 2). We did not find a high risk of methodological bias in the included studies (Table S3). The FMF preterm-PE prediction model includes the following predictors: maternal characteristics (age, height, weight, racial origin, method of conception, cigarette smoking, history of chronic hypertension and diabetes mellitus, history of systemic lupus erythematosus or antiphospholipid syndrome, nulliparity, parous without previous history of PE, parous with previous history of PE, family history of PE and interpregnancy interval), biophysical measurements (mean arterial pressure (MAP) and mean uterine artery pulsatility index (UtA-PI)) and biochemical measurements (serum PlGF and serum PAPP-A). The performance of the FMF preterm-PE prediction competing-risks model in external validation studies is summarized in Table S5.

Distribution of prognostic factors

A variety of prognostic factors were included in the any-, early- and late-onset models (Figure 2). The prognostic factors used in the external validation of any-, early- and late-onset PE prediction models included maternal demographic and clinical characteristics, biochemical markers and ultrasound markers. None of the included prediction models used the same type and number of prognostic factors across different studies. Overall, maternal demographic and clinical characteristics were the most frequently used prognostic factors. UtA-PI and PAPP-A were the most used biomarkers for external validation of original models across studies. Prepregnancy body mass index (BMI) (some studies reported using the first-trimester BMI taken as prepregnancy BMI),

Table 1 Characteristics of 28 studies performing external validation of models from 20 any-onset pre-eclampsia prediction studies

	External validation			
Derivation study	*Study*	*Country*	*Study design*	*Sample/events (events per predictor)*
Goetzinger (2010)[45]	Snell (2020)[30]	UK	IPD	8811/343 (69)
	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (17)
	Allen (2017)[27]	UK	Single-center	2186/52 (10)
Poon (2008)[33]	Snell (2020)[30]	UK	IPD	3257/102 (25)
	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (15)
	Allen (2017)[27]	UK	Single-center	2180/52 (13)
Plasencia (2007)[34]	Snell (2020)[30]	UK	IPD	3257/102 (20)
	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (17)
	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (15)
Baschat (2014)[46]	Snell (2020)[30]	UK	IPD	5257/287 (32)
	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (17)
Odibo (2011)[42]	Snell (2020)[30]	UK	IPD	59 892/1774 (443)
	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (21)
Sovio and Smith (2019)[44]	Brunelli (2020)[31]	Italy	Multicenter	7619/73 (8)
Wright (2015)[41]	Snell (2020)[30]	UK	IPD	1916/76 (8)
Odibo (2011)[42]	Snell (2020)[30]	UK	IPD	1145/28 (7)
Yu (2005)[43]	Snell (2020)[30]	UK	IPD	4212/273 (45)
Giguère (2015)[38]	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (43)
Goetzinger (2014)[39]	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (17)
Myatt (2013)[40]	Lamain-de Ruiter (2019)[29]	The Netherlands	Multicenter	3736/87 (29)
Macdonald-Wallis (2015)[49]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (10)
Kenny (2014)[51]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (25)
Direkvand-Moghadam (2013)[35]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (19)
North (2011)[50]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (8)
Seed (2011)[47]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (15)
Syngelaki (2011)[48]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (8)
Audibert (2010)[36]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (38)
Poon (2008)[37]	Meertens (2019)[28]	The Netherlands	Multicenter	2614/76 (13)

Only first author is given for each study. Data are given as *n*. Total number of included papers will not be equal to sum of articles presented in main text and tables, as some models were validated externally more than once. IPD, individual patient data.

ethnicity, history of chronic hypertension, history of PE, family history of PE, UtA-PI and PAPP-A were the most frequently used prognostic factors in external validation of original models (Figure 2).

Model discrimination and calibration

Any-onset models

For the any-onset PE models, each of the 28 external validations produced an AUC (Figure 3). Only 7/20 models had good (AUC > 0.70) discrimination performance on external validation and all seven models performed poorer than reported in the primary prediction-model studies. Eleven externally validated prediction models for any-onset PE used only maternal clinical characteristics as prognostic factors, while nine models used maternal clinical characteristics and biomarkers. Among the 11 prediction models that used only maternal clinical characteristics, five[33,34,37,48,49] models had good (AUC > 0.7) discrimination performance and were well calibrated on external validation. Only two[42,46] prediction models that used maternal clinical characteristics and biomarkers as prognostic factors had good discrimination performance (Figure 3). Thirteen any-onset PE externally validated prediction models used fewer than five (median value) prognostic factors, while the rest of the models used five or more prognostic factors. Among models including five or more prognostic factors, only four had good discrimination performance (Figure S1). Overall, only five models[33,34,42,48,49] had good discrimination and were well calibrated, whereas the rest of the models were likely overfitted, as shown by the poorer external validation results (Table S6).

Three any-onset PE prediction models were validated externally three times[33,34,45] and two were validated externally twice[42,46] in different studies[27–30]. Among the externally validated models that underwent validation more than once, no significant difference in model performance was found. The model developed by Poon *et al.*[33] was validated externally in three different studies[27,29,30], appeared well calibrated and had good discrimination performance (AUC, 0.71) in two studies[27,29]. The summary AUC was 0.71 (95% PI, 0.66–0.76). The model by Plasencia *et al.*[34] was validated externally in three different studies[28–30], two of which reported good discrimination, but variable calibration. The summary AUC was 0.73 (95% PI, 0.55–0.86). In addition, the model using maternal characteristics and biomarkers by Goetzinger *et al.*[45] was validated externally in three different studies[27,29,30], which reported poor discrimination performance, with point estimates of AUC ranging from 0.55 to 0.66. The summary AUC also indicated poor discrimination (AUC, 0.60 (95% PI, 0.048–0.98)) (Figure 3).

Early- and late-onset PE models

Five early-onset PE prediction models[46,59–62] were validated externally more than once in different studies. The prediction model developed by Scazzocchio *et al.*[59] was validated externally four times in different studies[26,27,30,32]. The AUC value ranged from 0.74 to 0.94 on external validation in four different studies, but the model was not well calibrated. A prediction model from the study by Parra-Cordero *et al.*[60] was validated externally twice in different studies and showed good discrimination performance on external validation, but was not well calibrated in the USA cohort[32]. Two prediction models[61,62] validated externally by the same study[27] had good model discrimination performance, but were not well calibrated on external validation. Overall, 10 early-onset PE

Table 2 Characteristics of studies performing external validation of Fetal Medicine Foundation (FMF) models for preterm pre-eclampsia (PE) prediction

					PE status		
Validation study	*Country*	*Center*	*Sample*	*Maternal age (years)*	*Any-onset*	*Preterm*	*Term*
Cuenca-Gómez (2023)[67]	Spain	Multicenter	10 110	34.2 (31.7–38.6)	230	72	158
Riishede (2023)[74]	Denmark	Multicenter	8156	30.8 (28.1–33.9)	303	55	248
Rolnik (2022)[75]	Australia	Multicenter	29 618	33.3 ± 4.3	455	132	323
Hu (2021)[76]	China	Multicenter	10 899	31.3 (28.1–33.6)	312	195	117
Goto (2021)[77]	Japan	Single	913	40 (34–47)	26	11	15
Prasad (2021)[78]	India	Single	1863	31.40 ± 3.63	59	25	34
Zwertbroek (2021)[79]	The Netherlands	Single	362	28 (25–32)	22	10	12
Rezende (2021)[66]*	Brazil	Single	1695	—	164	41	105
Chaemsaithong (2019)[80]	Multicountry†	Multicenter	10 935	34.76 (30.29–37.47)	224	73	151
Lobo (2019)[81]	Brazil	Single	617	30 (25–35)	34	18	16
Rezende (2019)[68]*	Brazil	Single	1531	30 (24–35)	120	26	94
Guizani (2018)[69]	Belgium	Single	3239	31.7 (21.0–43.8)	80	36	44
Tan (2018)[70]	Multicountry‡	Multicenter	61 174	32.1 (27.5–36.0)	1770	493	1277
Tan (2018)[71]	UK	Multicenter	16 451	31.5 (27.4–35.1)	439	135	304
O'Gorman (2017)[72]	Multicountry‡	Multicenter	8775	30.6 (26.0–34.7)	239	59	180
O'Gorman (2016)[73]	UK	Multicenter	35 948	31.5 (27.0–35.6)	1058	292	766

Only first author is given for each study. Data are given as *n*, median (interquartile range) or mean ± SD. *Externally validated using only maternal factors, mean arterial pressure and uterine artery pulsatility index. †Hong Kong SAR, China, Japan, Thailand, Taiwan, India and Singapore. ‡UK, Spain, Belgium, Italy and Greece.

prediction models had good discrimination performance, among which only three[41,60,62] were well calibrated on external validation (Table S7).

Among the 14 late-onset PE prediction models, only five had good discrimination performance (Table S8). Three late-onset PE prediction models[59–61] were validated externally more than once in different cohorts[26,27,30,32]; none of them had consistently good discrimination and calibration performance.

FMF competing-risks model prediction performance

Fourteen validation studies were included in meta-analyses, as two of the 16 did not fulfill criteria for meta-analysis. Four meta-analyses were performed for four different models, combining (i) maternal factors, triple test (MAP, UtA-PI, PlGF) and PAPP-A; (ii) maternal factors and triple test; (iii) maternal factors only; and (iv) maternal factors and MAP. The summary AUC in 12 studies[67,69–76,78,79,81] was 0.90 (95% PI, 0.77–0.96), using maternal factors, triple test and PAPP-A. The lowest AUC was 0.71 (95% CI, 0.51–0.90), reported by Zwertbroek *et al.*[79], and the greatest AUC was 0.96 (95% CI, 0.95–0.98), reported by Prasad *et al.*[78] (Figure 4a). The summary AUC for models with maternal factors and triple test[67,70,72,73,76,77,80] was 0.90 (95% PI, 0.76–0.96), with AUCs ranging from 0.83 (95% CI, 0.79–0.88)[76] to 0.95 (95% CI, 0.86–0.98)[77] (Figure 4b). Validation studies of the FMF competing-risks model using maternal factors only demonstrated a summary AUC of 0.77 (95% PI, 0.71–0.82), with AUCs ranging from 0.71 (95% CI, 0.65–0.77)[67] to 0.80 (95% CI,

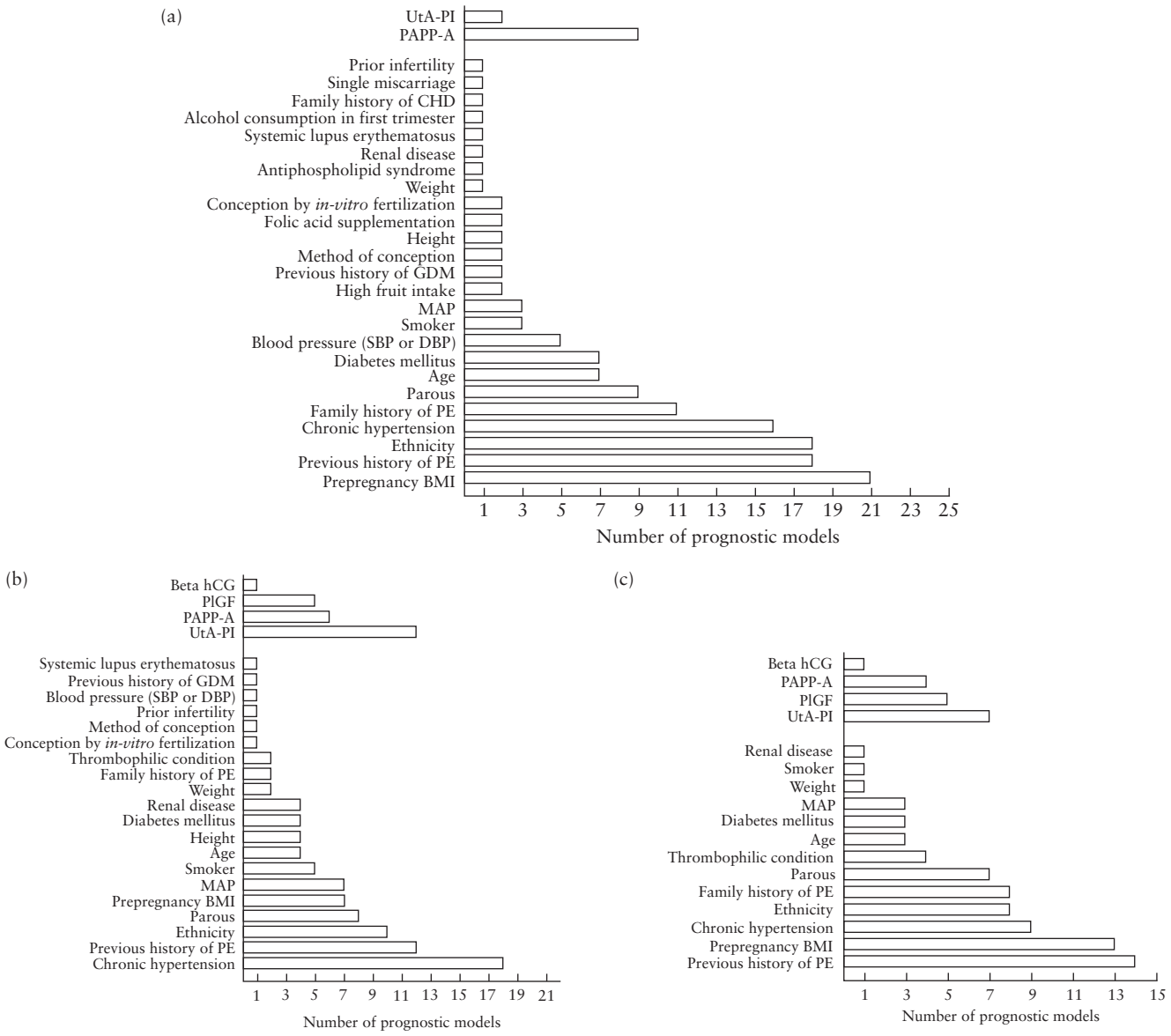

Figure 2 Distribution of prognostic factors across external validation model studies in models for any-onset pre-eclampsia (PE) (a), early-onset PE (b) and late-onset PE (c). BMI, body mass index; CHD, chronic heart disease; DBP, diastolic blood pressure; GDM, gestational diabetes mellitus; hCG, human chorionic gonadotropin; MAP, mean arterial pressure; PAPP-A, pregnancy-associated plasma protein-A; PlGF, placental growth factor; SBP, systolic blood pressure; UtA-PI, uterine artery pulsatility index.

0.77–0.82)[73] (Figure 4c). Validation studies including MAP with maternal factors had a summary AUC of 0.84 (95% PI, 0.82–0.85), with AUCs ranging from 0.80 (95% CI, 0.76–0.85)[67] to 0.91 (95% CI, 0.77–0.97)[77] (Figure 4d). The summary performance of the FMF competing-risks model for term PE prediction was 0.79 (95% PI, 0.69–0.87) (Figure S2). Only six studies[67,71,73–75,80] reported model calibration, with calibration slopes ranging from 0.95 to 1.05, indicating good calibration performance. The available prediction slopes and prediction intercepts are summarized in Table S5.

The summary discrimination performance was compared among the same externally validated models based on the predictor variables used. Five studies[67,70,72,73,76] evaluated the performance of the FMF competing-risks model based on maternal factors, triple test and PAPP-A *vs* maternal factors plus triple test. The summary AUC was 0.90 (95% PI, 0.76–0.97) for models including maternal factors, triple test and PAPP-A compared with 0.91 (95% PI, 0.73–0.97) for models including maternal factors and triple test (Figure S3a). Similarly, six studies[67,69,70,72,73,76] evaluated the performance of the

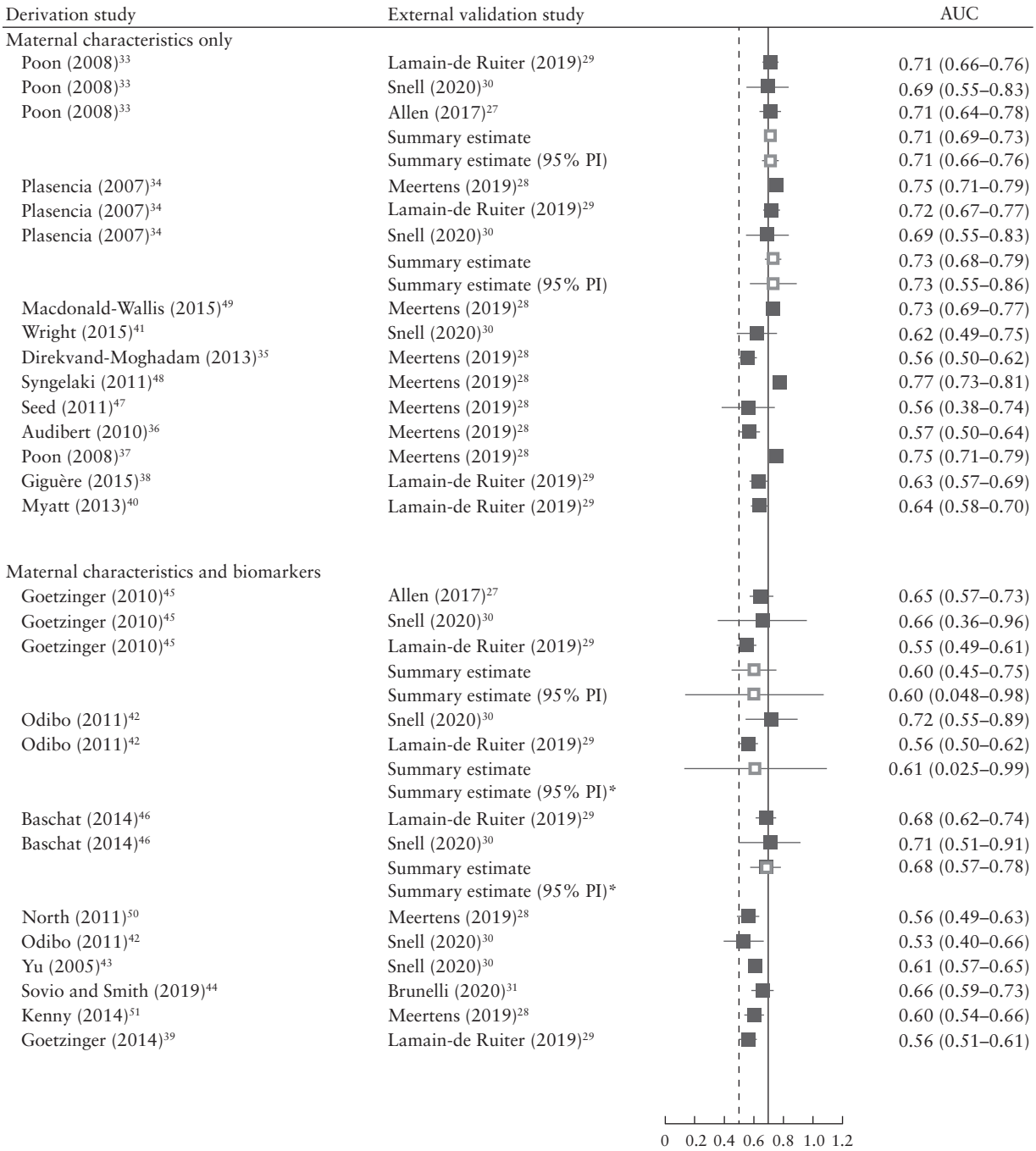

Figure 3 Forest plot showing area under receiver-operating-characteristics curve (AUC) of any-onset pre-eclampsia prediction models on external validation. Only first author is given for each study. Values in parentheses are 95% CI, unless indicated otherwise. *Prediction interval (PI) cannot be calculated.

FMF model based on maternal factors alone *vs* maternal factors plus MAP. The summary AUC of models that included only maternal factors was 0.77 (95% PI, 0.68–0.84), while models that included maternal factors plus MAP had an AUC of 0.84 (95% PI, 0.82–0.85) (Figure S3b).

DISCUSSION

PE is a predictable and preventable condition. Risk prediction models exist, but external validations of a risk prediction model and impact assessment studies should be completed before clinical implementation[82]. This systematic review found the triple-test FMF model demonstrates outstanding discrimination and calibration in preterm PE prediction, and is clearly the preferred model at present, having been validated externally in numerous settings. Other prediction models for PE have undergone more limited external validation. Apart from the FMF model, the majority of existing models had poor-to-good discrimination and poor calibration performance on external validation. Except for two studies[35,60], all other models were developed and validated externally in high-income settings. Overall, a wide variation of prognostic factors was used, with none having the same set of predictors. Interestingly, any-onset models incorporating only maternal characteristics performed similarly or slightly better than models incorporating additional biomarkers, with potential implications for low-resource and non-specialized settings.

The FMF model underwent 16 external validations, conducted in high- and middle-income countries. The FMF models were validated with comparable population age distribution, outcome definition and prognostic factors. Sixteen external validations may seem excessive; however, it demonstrates generalizability in new settings and clinicians and health administrations will be reassured regarding the model performance within similar populations and settings[83]. Only five (25%) any-onset PE models had undergone multiple external validations. These findings highlight the need for multiple external validations to determine if a model provides clinical utility and is generalizable[84–86], which may become more feasible as routine electronic health

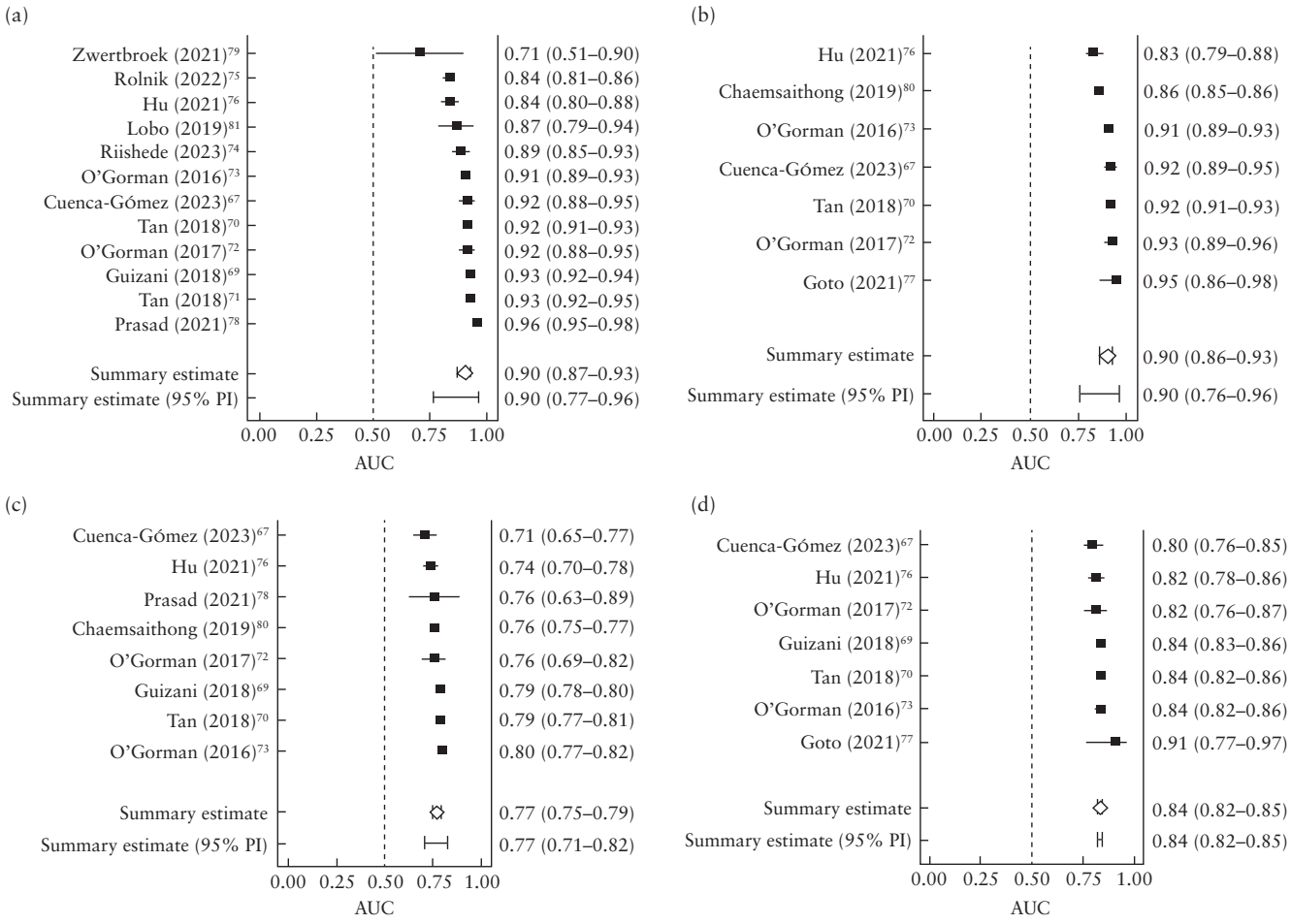

Figure 4 Forest plots showing area under receiver-operating-characteristics curve (AUC) of Fetal Medicine Foundation model for preterm pre-eclampsia prediction, including: (a) maternal factors plus triple test (mean arterial pressure, uterine artery pulsatility index and placental growth factor) plus pregnancy-associated plasma protein-A; (b) maternal factors plus triple test; (c) maternal factors only; and (d) maternal factors plus mean arterial pressure. Only first author is given for each study. Values in parentheses are 95% CI, unless indicated otherwise. PI, prediction interval.

data become available. The prediction performance of the FMF model was evaluated extensively in different countries[66–81]; however, limited information is available from low-resource countries, which highlights more validation needs.

Maternal demographic and clinical characteristics were the most frequently used prognostic factors, whereas UtA-PI and PAPP-A were frequent biomarkers in any-, early- and late-onset PE prediction. The FMF model[64] was originally developed in 2012 and later updated to include serum PlGF and PAPP-A biomarkers[65]. The summary FMF prediction showed outstanding prediction performance, while the addition of PAPP-A did not improve prediction performance. The summary discrimination performance with maternal factors alone was good, and the performance improved further with the addition of MAP (Figure 4). Even though incorporating specialized tests, such as UtA-PI and serum biomarkers, can improve prediction performance, these tests may not be available in low-resource settings and all public antenatal care clinics[87].

Examining performance of models that include only maternal factors *vs* models with additional biomarkers is useful whenever possible, as it can determine whether there are any benefits from having women undergo specialized tests during pregnancy. In this review, examining any-onset PE models landed itself to this examination. Externally validated any-onset model performance was assessed based on the type of prognostic factors included. Among the 11 any-onset models validated using maternal clinical characteristics only, five had good discrimination performance and were well calibrated. Nine models that also included biomarkers were validated, and most had poor discrimination performance (Figure 3). Overall, models that included maternal clinical characteristics only had similar or slightly better prediction performance than models with biomarkers in any-onset PE prediction. These are important findings, and future models incorporating biomarkers should routinely report model performance compared with a version of that model incorporating maternal characteristics only. If this is confirmed, this finding is beneficial because (i) it may encourage less expensive models, as biomarker extraction and identification can be expensive, (ii) prognostic factors should be readily available and less invasive, (iii) uptake of a decision tool might occur more readily, as healthcare settings are familiar with other clinically implemented prediction tools that are parsimonious and utilize easily available predictors[88–92] and (iv) these could be easily implemented in low-resource and non-specialized settings after model performance is optimized and clinical utility is established.

Given that maternal clinical characteristics are reported as highly important risk factors for PE[93,94], it was not unexpected that any-onset PE models including only maternal characteristics performed just as well, and perhaps slightly better, than models incorporating biomarkers. The risk of developing PE has been reported to increase 8-fold in women with prior PE, 7-fold in women with prepregnancy/early-gestational-age BMI > 30 kg/m^2, 5-fold in women with chronic hypertension and 3-fold in women with pregestational diabetes[93]. Women with a first-degree relative with PE have a 3-fold increased risk of developing PE[93,94]. Therefore, existing literature supports that maternal clinical characteristics are highly important factors for PE, which are easily available and less invasive.

To support the pipeline towards producing clinically useful prediction tools, there is a need to comprehensively validate feasible and cost-effective models for PE that would be applied in all clinical settings. A recent cost-effectiveness analysis in Germany comparing the FMF model *vs* standard screening showed that cost per case averted incurred an additional cost of €1400 per 100 women screened[95]. Other studies revealed that the FMF model is cost-effective compared with standard care[96,97]. The FMF model has been validated externally in numerous settings and appears to be relatively cost-effective; therefore, it is recommended that other models (with and without biomarkers) also follow this process.

Limitations of this systematic review relate to the limitations of the original models and include heterogeneity in prognostic predictors used in current model development. A lack of models in ethnically diverse and low- and middle-income settings is also a weakness. A further limitation is the unstable model performance of any-onset PE prediction on external validation.

In conclusion, this systematic review revealed that many existing externally validated models had limited discrimination and calibration performance in any-, early- and late-onset PE prediction. The triple-test FMF preterm PE model demonstrated outstanding discrimination and was well calibrated; however, the inclusion of specialized biomarkers may limit feasibility and implementation, especially in low-resource settings. The validated any-onset PE models with only maternal demographic and clinical characteristics had similar, or slightly better, discrimination performance compared to models with additional biomarkers; although this may change in the future as new biomarkers emerge. Most external validation studies were conducted in high-income countries, highlighting the need for future work in low-resource settings, utilizing standardized prognostic factors and harmonized data sources.

ACKNOWLEDGMENTS

S.A.T. is supported by a PhD scholarship funded by the Monash Graduate Scholarship (MGS) and the Monash International Tuition Scholarship (MITS). S.T. is a senior investigator for the National Institute for Health and Care Research (NIHR). H.J.T. is supported by a National Health and Medical Research Council (NHMRC) fellowship (2009326). Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

REFERENCES

1. Karrar SA, Hong PL. Preeclampsia. *StatPearls [Internet]*. StatPearls Publishing; 2022.
2. Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. *Eur J Obstet Gynecol Reprod Biol*. 2013;170(1):1-7. doi:10.1016/j.ejogrb.2013.05.005
3. Thornton C, Tooher J, Ogle R, von Dadelszen P, Makris A, Hennessy A. Benchmarking the hypertensive disorders of pregnancy. *Pregnancy Hypertens*. 2016;6(4):279-284. doi:10.1016/j.preghy.2016.04.009
4. Thornton C, Dahlen H, Korda A, Hennessy A. The incidence of preeclampsia and eclampsia and associated maternal mortality in Australia from population-linked datasets: 2000-2008. *Am J Obstet Gynecol*. 2013;208(6):476.e1-476.e5. doi:10.1016/j.ajog.2013.02.042
5. Brown MA, Magee LA, Kenny LC, et al. Hypertensive disorders of pregnancy: ISSHP classification, diagnosis, and management recommendations for international practice. *Hypertension*. 2018;72(1):24-43. doi:10.1161/HYPERTENSIONAHA.117.10803
6. Schneider S, Freerksen N, Maul H, Roehrig S, Fischer B, Hoeft B. Risk groups and maternal-neonatal complications of preeclampsia–current results from the national German Perinatal Quality Registry. *J Perinat Med*. 2011;39:257-265.
7. Williams D. Pre-eclampsia and long-term maternal health. *Obstet Med*. 2012;5(3):98-104.
8. Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol*. 2009;33(3):130-137. doi:10.1053/j.semperi.2009.02.010
9. Henderson JT, Vesco KK, Senger CA, Thomas RG, Redmond N. Aspirin use to prevent preeclampsia and related morbidity and mortality: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2021;326(12):1192-1206.
10. Duley L, Meher S, Hunter KE, Seidler AL, Askie LM. Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane Database Syst Rev*. 2019;10:CD004659.
11. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*. 2015;350(January):1-9. doi:10.1136/bmj.g7594
12. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380. doi:10.1136/bmj.e5595
13. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460. doi:10.1136/bmj.i6460
14. Ranstam J, Cook JA, Collins GS. Clinical prediction models. *Br J Surg*. 2016;103(13):1886. doi:10.1002/bjs.10242
15. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ*. 2023;381:e073538. doi:10.1136/bmj-2022-073538
16. Enticott J, Buck K, Shawyer F. Finding "hard to find" literature on hard to find groups: a novel technique to search grey literature on refugees and asylum seekers. *Int J Methods Psychiatr Res*. 2018;27(1):e1580. doi:10.1002/mpr.1580
17. Thong EP, Ghelani DP, Manoleehakul P, et al. Optimising cardiometabolic risk factors in pregnancy: a review of risk prediction models targeting gestational diabetes and hypertensive disorders. *J Cardiovasc Dev Dis*. 2022;9(2):55. doi:10.3390/jcdd9020055
18. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. *PloS One*. 2012;7(2):e32844. doi:10.1371/journal.pone.0032844
19. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org.
20. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58. doi:10.7326/M18-1376
21. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
22. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Vol 398. John Wiley & Sons; 2013.
23. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? *J Clin Epidemiol*. 2020;118:93-99. doi:10.1016/j.jclinepi.2019.09.016
24. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation and updating*. Vol 66. Springer; 2010. doi:10.1111/j.1541-0420.2010.01431.x
25. Debray T, Debray MT. Package 'metamisc.' 2021 https://r-forge.r-project.org/projects/metamisc/
26. Scazzocchio E, Crovetto F, Triunfo S, Gratacós E, Figueras F. Validation of a first-trimester screening model for pre-eclampsia in an unselected population. *Ultrasound Obstet Gynecol*. 2017;49(2):188-193. doi:10.1002/uog.15982
27. Allen RE, Zamora J, Arroyo-Manzano D, et al. External validation of preexisting first trimester preeclampsia prediction models. *Eur J Obstet Gynecol Reprod Biol*. 2017;217:119-125. doi:10.1016/j.ejogrb.2017.08.031
28. Meertens LJE, Scheepers HCJ, Van Kuijk SMJ, et al. External validation and clinical usefulness of first trimester prediction models for the risk of preeclampsia: a prospective cohort study. *Fetal Diagn Ther*. 2019;45(6):381-393. doi:10.1159/000490385
29. Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models for preeclampsia in a Dutch multicenter prospective cohort. *Hypertens Pregnancy*. 2019;38(2):78-88. doi:10.1080/10641955.2019.1584210
30. Snell KIE, Allotey J, Smuk M, et al. External validation of prognostic models predicting pre-eclampsia: individual participant data meta-analysis. *BMC Med*. 2020;18(1):302. doi:10.1186/s12916-020-01766-9
31. Brunelli E, Seidenari A, Germano C, et al. External validation of a simple risk score based on the ASPRE trial algorithm for preterm pre-eclampsia considering maternal characteristics in nulliparous pregnant women: a multicentre retrospective cohort study. *BJOG*. 2020;127(10):1210-1215. doi:10.1111/1471-0528.16246
32. 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(3):279-285.
33. Poon LCY, Kametas NA, Pandeva I, Valencia C, Nicolaides KH. Mean arterial pressure at 11+0 to 13+6 weeks in the prediction of preeclampsia. *Hypertension*. 2008;51(4 Part 2 Suppl):1027-1033. doi:10.1161/HYPERTENSIONAHA.107.104646
34. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol*. 2007;30(5):742-749. doi:10.1002/uog.5157
35. Direkvand-Moghadam A, Khosravi A, Sayehmiri K. Predictive factors for preeclampsia in pregnant women: a receiver operation character approach. *Arch Med Sci*. 2013;9(4):684-689. doi:10.5114/aoms.2013.36900
36. Audibert F, Boucoiran I, An N, et al. Screening for preeclampsia using first-trimester serum markers and uterine artery Doppler in nulliparous women. *Am J Obstet Gynecol*. 2010;203(4):383.e1-383.e8. doi:10.1016/j.ajog.2010.06.014
37. Poon LCY, Kametas N, Bonino S, Vercellotti E, Nicolaides KH. Urine albumin concentration and albumin-to-creatinine ratio at 11(+0) to 13(+6) weeks in the prediction of pre-eclampsia. *BJOG*. 2008;115(7):866-873. doi:10.1111/j.1471-0528.2007.01650.x
38. Giguère Y, Massé J, Thériault S, et al. Screening for pre-eclampsia early in pregnancy: performance of a multivariable model combining clinical characteristics and biochemical markers. *BJOG*. 2015;122(3):402-410. doi:10.1111/1471-0528.13050
39. Goetzinger KR, Tuuli MG, Cahill AG, Macones GA, Odibo AO. Development and validation of a risk factor scoring system for first-trimester prediction of preeclampsia. *Am J Perinatol*. 2014;31(12):1049-1056.
40. Myatt L, Ph D, Clifton RG, et al. First-trimester prediction of preeclampsia in nulliparous women at low-risk. *Obstet Gynecol*. 2013;119(6):1234-1242. doi:10.1097/AOG.0b013e3182571669.First-Trimester
41. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. *Am J Obstet Gynecol*. 2015;213(1):62.e1-62.e10. doi:10.1016/j.ajog.2015.02.018
42. Odibo AO, Zhong Y, Goetzinger KR, et al. First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. *Placenta*. 2011;32(8):598-602. doi:10.1016/j.placenta.2011.05.006
43. Yu CKH, Smith GCS, Papageorghiou AT, Cacho AM, Nicolaides KH; Fetal Medicine Foundation Second Trimester Screening Group. An integrated model for the prediction of preeclampsia using maternal factors and uterine artery Doppler velocimetry in unselected low-risk women. *Am J Obstet Gynecol*. 2005;193:429-436.
44. Sovio U, Smith GCS. Evaluation of a simple risk score to predict preterm pre-eclampsia using maternal characteristics: a prospective cohort study. *BJOG*. 2019;126(8):963-970. doi:10.1111/1471-0528.15664
45. Goetzinger KR, Singla A, Gerkowicz S, Dicke JM, Gray DL, Odibo AO. Predicting the risk of pre-eclampsia between 11 and 13 weeks' gestation by combining maternal characteristics and serum analytes, PAPP-A and free β-hCG. *Prenat Diagn*. 2010;30(12–13):1138-1142.
46. Baschat AA, Magder LS, Doyle LE, Atlas RO, Jenkins CB, Blitzer MG. Prediction of preeclampsia utilizing the first trimester screening examination. *Am J Obstet Gynecol*. 2014;211(5):514.e1-514.e7. doi:10.1016/j.ajog.2014.04.018
47. Seed PT, Chappell LC, Black MA, et al. Prediction of preeclampsia and delivery of small for gestational age babies based on a combination of clinical risk factors in high-risk women. *Hypertens Pregnancy*. 2011;30(1):58-73. doi:10.3109/10641955.2010.486460
48. Syngelaki A, Bredaki FE, Vaikousi E, Maiz N, Nicolaides KH. Body mass index at 11-13 weeks' gestation and pregnancy complications. *Fetal Diagn Ther*. 2011;30(4):250-265. doi:10.1159/000328083
49. Macdonald-Wallis C, Silverwood RJ, De Stavola BL, et al. Antenatal blood pressure for prediction of pre-eclampsia, preterm birth, and small for gestational age babies: development and validation in two general population cohorts. *BMJ*. 2015;351:351. doi:10.1136/bmj.h5948
50. North RA, McCowan LME, Dekker GA, et al. Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ*. 2011;342(7803):d1875. doi:10.1136/bmj.d1875
51. Kenny LC, Black MA, Poston L, et al. Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. *Hypertension*. 2014;64(3):644-652. doi:10.1161/HYPERTENSIONAHA.114.03578
52. Di Lorenzo G, Ceccarello M, Cecotti V, et al. First trimester maternal serum PlGF, free β-hCG, PAPP-A, PP-13, uterine artery Doppler and maternal history for the prediction of preeclampsia. *Placenta*. 2012;33(6):495-501. doi:10.1016/j.placenta.2012.03.003
53. Plasencia W, Maiz N, Poon L, Yu C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks and 21 + 0 to 24 + 6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol*. 2008;32(2):138-146. doi:10.1002/uog.5402
54. Poon LCY, Kametas NA, Maiz N, Akolekar R, Nicolaides KH. First-trimester prediction of hypertensive disorders in pregnancy. *Hypertension*. 2009;53(5):812-818. doi:10.1161/HYPERTENSIONAHA.108.127977
55. Crovetto F, Figueras F, Triunfo S, et al. First trimester screening for early and late preeclampsia based on maternal characteristics, biophysical parameters, and angiogenic factors. *Prenat Diagn*. 2015;35(2):183-191. doi:10.1002/pd.4519
56. Kuc S, Koster MPH, Franx A, Schielen PCJI, Visser GHA. Maternal characteristics, mean arterial pressure and serum markers in early prediction of preeclampsia. *PLoS One*. 2013;8(5):e63546. doi:10.1371/journal.pone.0063546

57. Poon LCY, Kametas NA, Chelemen T, Leal A, Nicolaides KH. Maternal risk factors for hypertensive disorders in pregnancy: a multivariate approach. *J Hum Hypertens*. 2010;24(2):104-110. doi:10.1038/jhh.2009.45
58. Caradeux J, Serra R, Nien JK, et al. First trimester prediction of early onset preeclampsia using demographic, clinical, and sonographic data: a cohort study. *Prenat Diagn*. 2013;33(8):732-736. doi:10.1002/pd.4113
59. Scazzocchio E, Figueras F, Crispi F, et al. Performance of a first-trimester screening of preeclampsia in a routine care low-risk setting. *Am J Obstet Gynecol*. 2013;208(3):203.e1-203.e10. doi:10.1016/j.ajog.2012.12.016
60. Parra-Cordero M, Rodrigo R, Barja P, et al. Prediction of early and late pre-eclampsia from maternal characteristics, uterine artery Doppler and markers of vasculogenesis during first trimester of pregnancy. *Ultrasound Obstet Gynecol*. 2013;41(5):538-544.
61. Poon LCY, Maiz N, Valencia C, Plasencia W, Nicolaides KH. First-trimester maternal serum pregnancy-associated plasma protein-A and pre-eclampsia. *Ultrasound Obstet Gynecol*. 2009;33(1):23-33. doi:10.1002/uog.6280
62. Poon LCY, Staboulidou I, Maiz N, Plasencia W, Nicolaides KH. Hypertensive disorders in pregnancy: screening by uterine artery Doppler at 11-13 weeks. *Ultrasound Obstet Gynecol*. 2009;34(2):142-148. doi:10.1002/uog.6452
63. Akolekar R, Zaragoza E, Poon LCY, Pepes S, Nicolaides KH. Maternal serum placental growth factor at 11 + 0 to 13 + 6 weeks of gestation in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol*. 2008;32(6):732-739. doi:10.1002/uog.6244
64. Wright D, Akolekar R, Syngelaki A, Poon LCY, Nicolaides KH. A competing risks model in early screening for preeclampsia. *Fetal Diagn Ther*. 2012;32(3):171-178. doi:10.1159/000338470
65. Akolekar R, Syngelaki A, Poon L, Wright D, Nicolaides KH. Competing risks model in early screening for preeclampsia by biophysical and biochemical markers. *Fetal Diagn Ther*. 2013;33(1):8-15. doi:10.1159/000341264
66. Rezende KB, Bornia RG, Rolnik DL, et al. External validation of first trimester combined screening for pre-eclampsia in Brazil: an observational study. *Pregnancy Hypertens*. 2021;26:110-115. doi:10.1016/j.preghy.2021.10.005
67. Cuenca-Gómez D, de Paco Matallana C, Rolle V, et al. Performance of first-trimester combined screening for preterm pre-eclampsia: findings from cohort of 10 110 pregnancies in Spain. *Ultrasound Obstet Gynecol*. 2023;62:522-530.
68. Rezende KBC, Cunha AJLAD, Amim Junior J, Bornia RG. External validation of the fetal medicine foundation algorithm for the prediction of preeclampsia in a Brazilian population. *Pregnancy Hypertens*. 2019;17:64-68. doi:10.1016/j.preghy.2019.05.006
69. Guizani M, Valsamis J, Dutemeyer V, et al. First-trimester combined multimarker prospective study for the detection of pregnancies at a high risk of developing preeclampsia using the fetal medicine foundation-algorithm. *Fetal Diagn Ther*. 2018;43(4):266-273. doi:10.1159/000477934
70. Tan MY, Syngelaki A, Poon LC, et al. Screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation. *Ultrasound Obstet Gynecol*. 2018;52(2):186-195.
71. Tan MY, Wright D, Syngelaki A, et al. Comparison of diagnostic accuracy of early screening for pre-eclampsia by NICE guidelines and a method combining maternal factors and biomarkers: results of SPREE. *Ultrasound Obstet Gynecol*. 2018;51(6):743-750.
72. O'Gorman N, Wright D, Poon LC, et al. Accuracy of competing-risks model in screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation. *Ultrasound Obstet Gynecol*. 2017;49(6):751-755. doi:10.1002/uog.17399
73. O'Gorman N, Wright D, Syngelaki A, et al. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation. *Am J Obstet Gynecol*. 2016;214(1):103.e1-103.e12. doi:10.1016/j.ajog.2015.08.034
74. Riishede I, Rode L, Sperling L, et al. Pre-eclampsia screening in Denmark (PRESIDE): national validation study. *Ultrasound Obstet Gynecol*. 2023;61:682-690.
75. Rolnik DL, Selvaratnam RJ, Wertaschnigg D, et al. Routine first trimester combined screening for preterm preeclampsia in Australia: a multicenter clinical implementation cohort study. *Int J Gynecol Obstet*. 2022;158(3):634-642. doi:10.1002/ijgo.14049
76. Hu J, Gao J, Liu J, et al. Prospective evaluation of first-trimester screening strategy for preterm pre-eclampsia and its clinical applicability in China. *Ultrasound Obstet Gynecol*. 2021;58(4):529-539.
77. Goto M, Koide K, Tokunaka M, et al. Accuracy of the FMF Bayes theorem-based model for predicting preeclampsia at 11–13 weeks of gestation in a Japanese population. *Hypertens Res*. 2021;44(6):685-691. doi:10.1038/s41440-020-00571-4
78. Prasad S, Sahota DS, Vanamail P, Sharma A, Arora S, Kaul A. Performance of Fetal Medicine Foundation algorithm for first trimester preeclampsia screening in an indigenous south Asian population. *BMC Pregnancy Childbirth*. 2021;21(1):1-7. doi:10.1186/s12884-021-04283-6
79. Zwertbroek EF, Groen H, Fontanella F, Maggio L, Marchi L, Bilardo CM. Performance of the FMF first-trimester preeclampsia-screening algorithm in a high-risk population in the Netherlands. *Fetal Diagn Ther*. 2021;48(2):103-111. doi:10.1159/000512335
80. Chaemsaithong P, Pooh RK, Zheng M, et al. Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population. *Am J Obstet Gynecol*. 2019;221(6):650.e1-650.e16. doi:10.1016/j.ajog.2019.09.041
81. Lobo GAR, Nowak PM, Panigassi AP, et al. Validation of fetal medicine foundation algorithm for prediction of pre-eclampsia in the first trimester in an unselected Brazilian population. *J Matern Neonatal Med*. 2019;32(2):286-292. doi:10.1080/14767058.2017.1378332
82. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698.
83. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med*. 2023;21(1):1-8. doi:10.1186/s12916-023-02779-w
84. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018
85. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007
86. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:338.
87. Al-Rubaie ZTA, Askie LM, Ray JG, Hudson HM, Lord SJ. The performance of risk prediction models for pre-eclampsia using routinely collected maternal characteristics and comparison with models that include specialised tests and with clinical guideline decision rules: a systematic review. *BJOG*. 2016;123(9):1441-1452.
88. Candido dos Reis FJ, Wishart GC, Dicks EM, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res*. 2017;19(1):1-13. doi:10.1186/s13058-017-0852-3
89. Rahmatinejad Z, Rahmatinejad F, Sezavar M, Tohidinezhad F, Abu-Hanna A, Eslami S. Internal validation and evaluation of the predictive performance of models based on the PRISM-3 (Pediatric Risk of Mortality) and PIM-3 (Pediatric Index of Mortality) scoring systems for predicting mortality in Pediatric Intensive Care Units (PICUs). *BMC Pediatr*. 2022;22(1):1-17. doi:10.1186/s12887-022-03228-y
90. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: Independent and external validation of an updated version of QRISK2. *BMJ*. 2012;345(7867):1-12. doi:10.1136/bmj.e4181
91. Vincent JL, De Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. *Crit Care Med*. 1998;26(11):1793-1800.
92. Sternbach GL. The Glasgow coma scale. *J Emerg Med*. 2000;19(1):67-71.
93. Bartsch E, Medcalf KE, Park AL, et al. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ*. 2016;353:i1753. doi:10.1136/bmj.i1753
94. Duckitt K, Harrington D. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *Br Med J*. 2005;330(7491):565-567. doi:10.1136/bmj.38380.674340.E0
95. Mewes JC, Lindenberg M, Vrijhoef HJM. Cost-effectiveness analysis of implementing screening on preterm pre-eclampsia at first trimester of pregnancy in Germany and Switzerland. *PLoS One*. 2022;17(6):e0270490. doi:10.1371/journal.pone.0270490
96. Dubon Garcia A, Devlieger R, Redekop K, Vandeweyer K, Verlohren S, Poon LC. Cost-utility of a first-trimester screening strategy versus the standard of care for nulliparous women to prevent pre-term pre-eclampsia in Belgium. *Pregnancy Hypertens*. 2021;25:219-224. doi:10.1016/j.preghy.2021.06.012
97. Park F, Deeming S, Bennett N, Hyett J. Cost-effectiveness analysis of a model of first-trimester prediction and prevention of preterm pre-eclampsia compared with usual care. *Ultrasound Obstet Gynecol*. 2021;58(5):688-697.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Appendix S1 Search strategy

Figure S1 Forest plot showing area under receiver-operating-characteristics curve reported from external validations of any-onset pre-eclampsia prediction models, according to whether number of prognostic factors included was $<$ or ≥ 5. Only first author is given for each study.

Figure S2 Forest plot showing area under receiver-operating-characteristics curve of Fetal Medicine Foundation model for term pre-eclampsia prediction. Only first author is given for each study.

Figure S3 Forest plots showing area under receiver-operating-characteristics curve of Fetal Medicine Foundation model for preterm pre-eclampsia prediction, including: (a) maternal factors plus triple test plus pregnancy-associated plasma protein-A *vs* maternal factors plus triple test; and (b) maternal factors only *vs* maternal factors plus mean arterial pressure. Only first author is given for each study.

Table S1 Characteristics of externally validated early-onset pre-eclampsia prognostic models

Table S2 Characteristics of externally validated late-onset pre-eclampsia prognostic models

Table S3 Risk of bias and applicability of included studies

Table S4 Details of externally validated any-onset, early-onset and late-onset pre-eclampsia prognostic models

Table S5 Performance/accuracy of Fetal Medicine Foundation preterm pre-eclampsia competing-risks models in external validation studies

Tables S6, S7 and S8 Model discrimination and calibration of any-onset (Table S6), early-onset (Table S7) and late-onset (Table S8) pre-eclampsia prognostic models on external validation