External Validation and Clinical Usefulness of First Trimester Prediction Models for the Risk of Preeclampsia: A Prospective Cohort Study

Linda J.E. Meertens^a Hubertina C.J. Scheepers^b Sander M.J. van Kuijk^c
Robert Aardenburg^d Ivo M.A. van Dooren^e Josje Langenveld^d
Annemieke M. van Wijck^f Iris M. Zwaan^g Marc E.A. Spaanderman^b
Luc J.M. Smits^a

^aDepartment of Epidemiology, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands; ^bDepartment of Obstetrics and Gynaecology, School for Oncology and Developmental Biology (GROW), Maastricht University Medical Center, Maastricht, The Netherlands; ^cDepartment of Clinical Epidemiology and Medical Technology Assessment (KEMTA), Maastricht University Medical Center, Maastricht, The Netherlands; ^dDepartment of Obstetrics and Gynaecology, Zuyderland Medical Center, Heerlen, The Netherlands; ^eDepartment of Obstetrics and Gynaecology, Sint Jans Gasthuis Weert, Weert, The Netherlands; ^fDepartment of Obstetrics and Gynaecology, VieCuri Medical Center, Venlo, The Netherlands; ^gDepartment of Obstetrics and Gynaecology, Laurentius Hospital, Roermond, The Netherlands

Keywords

Clinical usefulness · Decision curve analysis · External validation · First trimester · Prediction model · Preeclampsia · Risk assessment

Abstract

Introduction: This study assessed the external validity of all published first trimester prediction models for the risk of preeclampsia (PE) based on routinely collected maternal predictors. Moreover, the potential utility of the best-performing models in clinical practice was evaluated. **Material and Methods:** Ten prediction models were systematically selected from the literature. We performed a multicenter prospective cohort study in the Netherlands between July 1, 2013, and December 31, 2015. Eligible pregnant women completed a web-based questionnaire before 16 weeks'

gestation. The outcome PE was established using postpartum questionnaires and medical records. Predictive performance of each model was assessed by means of discrimination (c-statistic) and a calibration plot. Clinical usefulness was evaluated by means of decision curve analysis and by calculating the potential impact at different risk thresholds. **Results:** The validation cohort contained 2,614 women of whom 76 developed PE (2.9%). Five models showed moderate discriminative performance with c-statistics ranging from 0.73 to 0.77. Adequate calibration was obtained after refitting. The best models were clinically useful over a small range of predicted probabilities. **Discussion:** Five of the ten included first trimester prediction models for PE showed moderate predictive performance. The best models may provide more benefit compared to risk selection as used in current guidelines.

© 2018 The Author(s)
Published by S. Karger AG, Basel

Introduction

Preeclampsia (PE) is still a leading cause of maternal and neonatal morbidity and mortality worldwide. Although management of PE has improved, the only available curative treatment is delivery of the placenta. Further reduction of the burden of disease may be achieved through prevention [1, 2].

Calcium supplementation and low-dose aspirin are well-researched interventions that reduce the risk of developing PE. A daily intake of at least 1 g of calcium lowers the risk of PE in women with a low dietary calcium intake and women at increased risk of hypertensive disorders [3]. Low-dose aspirin safely reduces the occurrence of PE by 10–43% among high-risk women, and also reduces related adverse outcomes (e.g., preterm birth, small for gestational age) [4–10]. A recent meta-analysis, including the large Aspirin for Evidence-Based Preeclampsia Prevention (ASPREE) trial, demonstrated that aspirin mainly reduces the risk of preterm PE with a reported risk reduction of 38–67% [11, 12].

Efficiency of preventive strategies may improve if they are specifically aimed at women that are most likely to benefit from them. Prediction models may help identifying women at increased risk of PE early on in pregnancy, a crucial period at which preventive interventions are most effective.

A considerable number of first trimester prediction models for PE have been published, containing predictors ranging from maternal risk factors, blood pressure measurement, to more specialized tests such as uterine artery Doppler ultrasound and biomarkers (e.g., pregnancy-associated plasma protein-A, placental growth factor). Although specialized tests have been reported to improve predictive performance, a drawback is that these tests are not routinely performed or readily available in general antenatal settings, and generate substantial additional costs [13, 14].

Only a minority of the models have been validated in independent populations [14]. Prediction models usually perform less well in populations other than the ones in which they were developed [15]. External validation of models is necessary in order to gain insight into their robustness across populations and domains [15, 16].

In this paper, we report the results of an independent validation study of all published first trimester prediction models based on routinely collected maternal characteristics for the risk of PE, in a cohort of 2,614 Dutch pregnant women. Next to this, we evaluated clinical usefulness of the best performing models by means of decision

curve analysis and by estimating clinical impact at different risk thresholds and different management scenarios, and compared it with the performance of current international recommendations.

Material and Methods

Selection Prediction Models

We searched PubMed to identify first trimester prediction models for development of PE. The search was first performed in April 2013, before finalizing study questionnaires, and then updated until April 2017. The search strategy retrieved 1,361 citations (see online suppl. File 1; for all online suppl. material, see www.karger.com/doi/10.1159/000490385). L.M. screened all citations and, together with L.S., assessed eligibility of full text articles. In case of disagreement, a third reviewer (H.S.) was available.

Fifty articles described prediction models fulfilling the eligibility criteria, which were: (1) article presented the development of a prediction model or an update of a previously developed model, (2) endpoint of the model was the risk of developing PE, (3) model contained multiple predictors, (4) model was based on weighted risk predictors, (5) predictors were routinely collected in Dutch obstetric practice (maternal characteristics, anthropometric measures, or blood pressure measurements), and (6) predictors were available before 16⁺⁰ weeks of gestation. Citation lists of relevant articles were checked and yielded 9 additional eligible articles. L.M. contacted authors if the model intercept, estimates or definition of predictors were not provided in the article. If regression coefficients were nevertheless not obtained, the model was excluded from the validation analyses. We excluded 36 articles for the following reasons: algorithm not available ($n = 19$), model already published in another article ($n = 17$). The 23 included articles described 33 prediction models, 10 of which were aimed at predicting overall PE, 12 models at early PE, 6 models at late PE, 2 models at severe PE, and 4 models at predicting gestational age at delivery with PE. We only validated the models predicting any PE as the number of women with early onset PE (e.g., delivery <34 weeks of gestational age) was too low in our validation cohort ($n = 5$). None of the models were used in daily care during the study period.

The characteristics of included models are summarized in Table 1 (see Supplementary Table 1 for a comprehensive overview) [17–26]. The models were published between 2005 and 2015 in the UK, Iran, Canada, and Australia by 6 different research groups. Three models were developed for nulliparous women [18, 21, 23]. References of excluded studies are described in online Supplementary File 1.

Validation Cohort

We performed a multicenter prospective cohort study (Expect Study I) in the south-eastern part of the Netherlands (province of Limburg) [27]. Pregnant women less than 16 weeks of pregnancy and aged 18 years or older were recruited between July 1, 2013 and January 1, 2015 in 36 midwifery practices (primary care) and 6 hospitals (secondary and tertiary care). Follow-up took place until December 31, 2015. We excluded pregnancies that terminated before 24 weeks of gestation or for which the outcome was not available. Eligible pregnant women were asked to complete a web-based questionnaire before 16 weeks of gestation (pregnancy question-

Table 1. Characteristics of included prediction models for external validation

First author [Ref.], year	Country	Study design	PE cases, <i>n</i> /total	Predictors	Internally validated	Other external validation studies, AUROC (CI)
Macdonald-Wallis [17], 2015	UK	Prospective cohort	317/12,996 (2.4%)	Age, BMI, CH, DM, ethnicity, height, MAP, parity, prior GDM, prior PIH, smoking	No	UK: 0.79 (0.73–0.85) [17]
Kenny [18], 2014	SCOPE study	Prospective cohort	196/3,747 (4.9%)	BMI, fam. history PE, fruit intake, MAP	Yes	–
Direkvand-Moghadam [19], 2013	Iran	Cohort	58/610 (9.5%)	CH, prior infertility, prior PE	No	–
Syngelaki [20], 2011	UK	Prospective cohort	NR/41,577 (NR)	Age, BMI, CH, DM, ethnicity, fam. history PE, method conception, parity, prior PE, smoking	No	–
North [21], 2011	SCOPE study	Prospective cohort	186/3,529 (5.3%)	Age, alcohol, BMI, fam. history CHD, fam history PE, fruit intake, MAP, single miscarriage, smoking, time to conceive, vaginal bleeding, woman's birth weight	Yes	–
Seed [22], 2011	UK	RCT	190/1,121 (17%)	CH, DBP, ethnicity, folic acid supplementation, prior PE, SBP	Yes	–
Audibert [23], 2010	Canada	Prospective cohort	40/893 (6.2%)	BMI, preexisting medical condition	No	–
Poon [24], 2008	UK	Prospective cohort	51/2,486 (2.0%)	Age, BMI, ethnicity, fam. history PE, parity, prior PE	No	–
Poon [25], 2008	UK	Prospective cohort	104/5,193 (2.3%)	BMI, ethnicity, fam. history PE, parity, prior PE, MAP	No	UK: 0.71 (0.64–0.78) [40]
Plasencia [26], 2007	UK	Prospective cohort	107/6,015 (1.8%)	BMI, ethnicity, fam. history PE, parity, prior PE	No	Italy: 0.72 (0.62–0.82) [41] Spain: early PE 0.74 (0.60–0.89), late PE 0.65 (0.49–0.80) [42]

AUROC, area under the receiver operating characteristic curve; BMI, body mass index; CH, chronic hypertension; CHD, coronary heart disease; CI, confidence interval; DBP, diastolic blood pressure; DM, diabetes mellitus; GDM, gestational diabetes mellitus; MAP, mean arterial pressure; NR, not reported; PE, preeclampsia; PIH, pregnancy-induced hypertension; RCT, randomized controlled trial; SBP, systolic blood pressure; SCOPE study, multicenter study Australia, New Zealand, and UK.

naire) and at 6 weeks after the due date (postpartum questionnaire). The web-based questionnaires could be accessed through the study website by use of a personal login code provided with the study information. Paper-and-pencil questionnaires were made available upon request. Automatic reminders were sent out in case of incompleteness or nonresponse. Medical records and letters of discharge about the pregnancy and birth were requested from caregivers.

The Medical Ethical Committee of the Maastricht University Medical Center evaluated the study protocol and declared that no ethical approval was necessary (MEC 13-4-053). All participating women gave informed consent through the Internet.

Assessment of Predictors

Predictors were assessed by means of the pregnancy questionnaire. Systolic and diastolic blood pressure (mm Hg) and heart rate (beats per minute) were measured before 16 weeks of gestation by

the gynecologist or midwife. The results of these measurements were subsequently self-reported in the pregnancy questionnaire. The predictors were defined as described in the original articles. For the predictor woman's birth weight, a proxy variable was made as we measured this variable in categories instead of as a continuous variable. Since no measurements were available for crown-rump length at blood pressure measurement, we made a proxy variable based on gestational age [28]. A detailed description about definition and assessment of the predictors is provided in online Supplementary Table 2.

Assessment of PE

The Dutch national guideline defines PE as systolic blood pressure ≥ 140 mm Hg and/or diastolic blood pressure ≥ 90 mm Hg (Korotkoff V) after 20 weeks gestation, measured twice in a previously normotensive woman, accompanied by proteinuria (at least

Fig. 1. Flowchart of the validation cohort.

300 mg protein in a 24-h urine collection) [29, 30]. The outcome was obtained from a combination of the medical record and postpartum questionnaire. We checked whether the medical record stated a diagnosis of PE or a combination of the field codes proteinuria and highest diastolic blood pressure measurement ≥ 90 mm Hg. Women were asked to report any occurrence of PE in the postpartum questionnaire. PE was defined as present when the diagnosis was confirmed by both sources. In the absence of the postpartum questionnaire ($n = 435$), the medical record was used as a gold standard and vice versa ($n = 16$). In case of a discrepancy between the pregnancy questionnaire and medical record, we contacted the obstetric caregiver for final decision.

Statistical Analysis

There is no generally accepted rule for the required sample size for external validation studies of prediction models. Vergouwe et al. [31] state that at least 100 events and 100 nonevents are necessary in order to be able to detect relevant differences between model performance in the derivation set and the validation set. Expecting that around 4% of the pregnancies would be affected by PE and accounting for incomplete records (10%), recruitment of 2,750 women was necessary.

Baseline characteristics were analyzed in the validation cohort and presented as means $\pm$ standard deviation for continuous variables and an absolute value with percentage for categorical variables. We imputed the missing values using stochastic regression imputation using predictive mean matching as the imputation model [32]. We evaluated the relatedness of the predictors and outcome between the validation cohort and the included original cohorts (case-mix).

For each subject in our validation cohort, we computed probabilities of developing PE on the basis of all included models. The algorithms of the included models are presented in online Supplementary Table 3. Predictive performance of the models was assessed by means of discrimination and calibration. Discrimination is the models' ability to distinguish between women who will develop PE and those who will not. We quantified discrimination by calculating the c-statistic (area under the receiver operating characteristic [ROC] curve, or AUROC) with 95% confidence interval. Calibration is the degree of agreement between the predicted probabilities and observed outcomes, and was examined by construct-

ing calibration plots. Perfect predictions should be on the 45° line with an intercept of 0 (calibration in the large) and a slope of 1 [33]. Calibration in the large compares the mean predicted risk with the observed proportion of PE cases, and indicates whether predictions are systematically too low or too high [33]. Calibration slope refers to the average strength of the predictor effects [15]. At validation, the slope is often smaller than one, reflecting overfitting of the model (low predictions too low and high predictions too high) [33]. We divided the women into 10 groups of roughly equal size with similar predicted risk. In case of miscalibration, we recalibrated the prediction model by adjusting the intercept and slope using the linear predictor as the only covariable [15]. This recalibration method has no effect on the discriminative performance of the model as the spread and ranking of the predictions remain the same.

We used the validation cohort with our own defined eligibility criteria for the analyses of each model. However, the original models were developed in different pregnant populations with their own in- and exclusion criteria. We additionally assessed the discriminative performance for each original study according to their specific eligibility criteria. Furthermore, a subgroup analysis was performed among nulliparous women, as previous PE was a strong predictor in most models.

Lastly, we determined the potential impact of the (recalibrated) models in clinical practice. Decision curve analysis was carried out for the best performing models as it provides a first impression of the clinical usefulness of the models. Decision curve analysis assesses net benefit (net proportion of true positives) of the models over a range of threshold probabilities compared to the scenarios of classifying all and no women as high risk of developing PE [34]. For the model with the highest net benefit, we evaluated the potential clinical impact of the model at different risk thresholds and different management strategies. We calculated incidence reduction of PE and number needed to prevent one case of PE using theoretical preventive interventions with relative risk reductions of 10, 20, and 40%. We also compared the model performance to current guidelines, the NICE and ACOG criteria [35–37]. As the ACOG criteria do not specifically define the number of moderate risk factors to identify women at high risk for PE, we defined ≥ 2 moderate risk factors as high risk [36]. Low socioeconomic status was defined as education below the level of tertiary education.

Table 2. Baseline characteristics of the validation cohort (Expect Study I)

Characteristics	Missing values, <i>n</i> (%)	Observed validation cohort (Expect Study I) ^a		
		overall (<i>n</i> = 2,614)	preeclampsia (<i>n</i> = 76)	no preeclampsia (<i>n</i> = 2,538)
Age, years	0 (0.0)	30.2 (3.9)	29.2 (4.1)	30.2 (3.9)
Ethnicity	0 (0.0)			
Caucasian		2,533 (96.9)	73 (96.1)	2,460 (96.9)
Afro-Caribbean		3 (0.1)	0 (0.0)	3 (0.1)
Indian, Pakistani or Bangladeshi		4 (0.2)	0 (0.0)	4 (0.2)
East Asian		4 (0.2)	0 (0.0)	4 (0.2)
Other Asian		12 (0.4)	1 (1.3)	11 (0.4)
Hispanic		11 (0.4)	1 (1.3)	10 (0.4)
Mixed		47 (1.8)	1 (1.3)	46 (1.8)
Tertiary education	3 (0.1)	1,420 (54.3)	35 (46.1)	1,385 (54.6)
Height, cm	3 (0.1)	168.8 (6.4)	168.7 (6.2)	168.8 (6.4)
Weight, kg	5 (0.2)	68.9 (13.0)	70.4 (12.8)	68.8 (13.0)
Body mass index	5 (0.2)	24.2 (4.3)	24.7 (3.9)	24.1 (4.3)
Smoking				
Ever <16 weeks' gestation	1 (0.0)	318 (12.2)	8 (10.5)	310 (12.2)
Current (at completion questionnaire)	1 (0.0)	157 (6.0)	2 (2.6)	155 (6.1)
Number cigarettes/day	1 (0.0)	6.8 (3.9)	11 (5.7)	6.8 (3.9)
Alcohol consumption ever <16 weeks' gestation	3 (0.1)	479 (18.3)	10 (13.2)	469 (18.5)
Diabetes mellitus	0 (0.0)			
Type 1		11 (0.4)	1 (1.3)	10 (0.4)
Type 2		9 (0.3)	1 (1.3)	8 (0.3)
Other		1 (0.0)	0 (0.0)	1 (0.0)
History of chronic hypertension	0 (0.0)	28 (1.1)	2 (2.6)	26 (1.0)
Chronic renal disease	0 (0.0)	5 (0.2)	0 (0.0)	5 (0.2)
Systemic lupus erythematosus	0 (0.0)	3 (0.1)	0 (0.0)	3 (0.1)
Antiphospholipid syndrome	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Hyperthyroid	0 (0.0)	11 (0.4)	0 (0.0)	11 (0.4)
High fruit intake ≥3/day	1 (0.0)	67 (2.6)	0 (0.0)	67 (2.6)
Family history of coronary heart disease (father)	1 (0.0)	439 (16.8)	11 (14.5)	428 (16.9)
Family history of preeclampsia				
Mother	2 (0.1)	91 (3.5)	5 (6.6)	86 (3.4)
Sister	4 (0.2)	41 (1.6)	1 (1.3)	40 (1.6)
Mother or sister	6 (0.2)	130 (5.0)	6 (7.9)	124 (4.9)
Woman's birth weight, g	319 (12.2)	3,174.1 (475.6)	3,062.5 (556.3)	3,177.3 (472.8)
Parity	0 (0.0)			
Nulliparous		1,326 (50.7)	66 (86.8)	1,260 (49.6)
Primiparous		1,026 (39.3)	10 (13.2)	1,016 (40.0)
Multiparous		262 (10.0)	0 (0.0)	262 (10.4)
Time to conceive ≥12 months	9 (0.3)	357 (13.7)	22 (28.9)	335 (13.2)
Conception	0 (0.0)			
Spontaneous		2,440 (93.3)	63 (82.9)	2,377 (93.7)
Ovulation induction		93 (3.6)	4 (5.3)	89 (3.5)
IVF/ICSI		81 (3.1)	9 (11.8)	72 (2.8)
Previous single miscarriage ≤10 weeks' gestation, same father	0 (0.0)	395 (15.1)	7 (9.2)	388 (15.3)
Folic acid supplementation at completion questionnaire	3 (0.1)	2,226 (85.2)	71 (93.4)	2,155 (84.9)
Vaginal bleeding ≥5 days	0 (0.0)	102 (3.9)	3 (3.9)	99 (3.9)
History of pregnancy-induced hypertension	18 (0.7)	115 (4.4)	3 (3.9)	112 (4.4)
History of preeclampsia	18 (0.7)	72 (2.8)	5 (6.6)	67 (2.6)
History of gestational diabetes mellitus	19 (0.7)	15 (0.6)	0 (0.0)	15 (0.6)
Systolic blood pressure, mm Hg	260 (9.9)	114.4 (12.5)	116.9 (13.7)	114.4 (12.4)
Diastolic blood pressure, mm Hg	270 (10.3)	67.6 (8.5)	69.7 (8.5)	67.6 (8.5)
Mean arterial pressure, mm Hg	270 (10.3)	83.2 (8.8)	85.4 (9.3)	83.2 (8.8)
Crown-rump length, mm	270 (10.3)	33.8 (23.5)	30.4 (21.1)	33.9 (23.6)

ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilization. ^a Original data (not imputed) presented as mean (SD) or absolute *n* (%).

Table 3. Discrimination of selected prediction models for preeclampsia

First author [Ref.], year	AUROC (95% CI) Original publication	AUROC (95% CI) Validation cohort (<i>n</i> = 2,614)	AUROC (95% CI) Validation cohort Definition original publication
Macdonald-Wallis [17], 2015	0.77 (0.75–0.80) <i>External validation:</i> 0.79 (0.73–0.85)	0.73 (0.68–0.77)	0.73 (0.68–0.77)
Kenny [18], 2014	0.73 (0.69–0.76) <i>Internal validation:</i> 0.64 (0.59–0.69)	0.60 (0.54–0.67)	0.60 (0.53–0.66)
Direkvand-Moghadam [19], 2013	0.67 (NR)	0.56 (0.50–0.62)	0.56 (0.50–0.62)
Syngelaki [20], 2011	NR	0.77 (0.72–0.81)	0.77 (0.72–0.81)
North [21], 2011	0.76 (NR) <i>Internal validation:</i> 0.71 (NR)	0.58 (0.51–0.64)	0.56 (0.48–0.63)
Seed [22], 2011	0.70 (NR) <i>Internal validation:</i> 0.66 (NR)	0.52 (0.46–0.59)	0.56 (0.38–0.74)
Audibert [23], 2010	0.75 (0.67–0.83)	0.56 (0.49–0.62)	0.57 (0.50–0.64)
Poon [24], 2008	0.85 (0.84–0.87)	0.74 (0.69–0.78)	0.75 (0.70–0.79)
Poon [25], 2008	<i>A priori risk (A):</i> 0.80 (NR) <i>A priori risk + MAP (B):</i> 0.85 (NR)	<i>A priori risk (A):</i> 0.74 (0.70–0.79) <i>A priori risk + MAP (B):</i> 0.63 (0.57–0.69)	<i>A priori risk (A):</i> 0.75 (0.70–0.79) <i>A priori risk + MAP (B):</i> 0.63 (0.57–0.70)
Plasencia [26], 2007	0.806 (0.80–0.82)	0.74 (0.70–0.78)	0.75 (0.71–0.79)

AUROC, area under the receiver operating characteristic curve; NR, not reported; MAP, mean arterial pressure.

All statistical analyses were conducted with IBM SPSS statistics version 23 (SPSS, Chicago, IL, USA) and R version 3.2.3, packages rms, pROC, and DecisionCurve.

Results

Validation Cohort

We included a total of 2,614 women in the validation cohort (Fig. 1). Table 2 shows the characteristics of the validation cohort in the observed data. For most predictors, the percentage of missing data was less than one. However, woman's own birth weight, systolic blood pressure, and diastolic blood pressure were missing in 12.2, 9.9, and 10.3% of women, respectively. An overview of the number of missing values per predictor and the characteristics of both the complete cases and the validation cohort after imputation is provided in online Supplemen-

tary Table 4. The pregnancy was complicated by PE in 76 women (2.9%).

We also described the characteristics of the development samples and compared these with the current validation sample (see online suppl. Table 5). The occurrence of the outcome was considerably higher in the development cohorts of Direkvand-Moghadam et al. [19] and Seed et al. [22]. The case-mix of Direkvand-Moghadam et al. [19], Seed et al. [22], and North et al. [21] substantially differed. In contrast to most development cohorts, our validation cohort had a low incidence of non-Caucasian ethnicity. Kenny et al. [18] and Audibert et al. [23] reported no predictor characteristics, and Syngelaki et al. [20] did not describe the number of cases.

Predictive Performances

The discriminative performance of the models is presented in Table 3. The AUROC decreased in five out of

Fig. 2. Calibration plots of externally validated first trimester prediction models for preeclampsia (PE). The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). Asterisks denote different limits on x and y axes.

ten models to below or equal to 0.60. The model of Syngelaki et al. [20] showed the best discriminative ability (AUROC 0.77). The models of Macdonald-Wallis et al. [17] and Kenny et al. [18] remained most stable after external validation (0.04 decline in AUROC). The ROC curves are available in online Supplementary Figure 1. The discriminative performances did not change much when validating the models in the cohort according to their own eligibility criteria.

Calibration plots of the models are provided in Figure 2. The distribution of the predicted probabilities was closely clustered around the mean for most models. Models consistently overestimated the risk of PE, except for the model of Poon et al. [24], Poon et al. [25], and Plasencia et al. [26]. All models were overfitted (e.g., the calibration slope < 1). The model of Poon et al. [25] was best calibrated to our population (slope 0.92). The model of Syngelaki et al. [20] had a slope of 1.1. Recalibration of the

models showed closer fitting to the ideal calibration line (Fig. 3).

Secondary analysis, in which the models were only applied to nulliparous women ($n = 1,326$), showed that the discriminative performance decreased considerably for all models with AUROC ranging from 0.51 to 0.61 (see online suppl. Fig. 2). The models specifically developed for nulliparous women (Kenny et al. [18], North et al. [21], and Audibert et al. [23]) did not perform better than the other models. The calibration plots are available in online Supplementary Figure 3. The model of Syngelaki et al. [20] had a slope of 0.59.

Clinical Usefulness

Figure 4 shows the net benefit of the five best performing models over a range of risk thresholds. The models were clinically useful, compared with classifying all or no women as high-risk, over a small range of probability thresholds (Macdonald-Wallis et al. [17] 0.50–5.07%, Syngelaki et al. [20] 0.52–5.98%, Poon et al. [24] 0.52–6.08%, Poon et al. [25] 0.52–6.44%, Plasencia et al. [26] 0.68–6.31%). Although the curves differed only slightly, the model of Syngelaki et al. [20] had the highest net ben-

efit for almost the complete range of clinical useful threshold probabilities.

The prognostic measures of the model of Syngelaki et al. [20] and the potential effect of the interventions on PE incidence reduction for six different risk thresholds are presented in Table 4. A low risk threshold ensures that almost no cases will be missed (high sensitivity), but many women will be treated unnecessarily. Conversely, a high specificity leads to a low proportion of false positives. However, the effect of the intervention will be smaller since more women prone to develop PE will be considered as low risk. The number needed to prevent one case of PE is dependent on the efficacy of the intervention, but less so of risk threshold.

We compared the model of Syngelaki et al. [20] with current guidelines. When applying the NICE criteria to our population, 8.5% of the women were considered as high risk with a sensitivity of 23.7% and a specificity of 91.9%. The ACOG criteria indicated 43.6% of our women as high-risk with a sensitivity of 68.4% and specificity of 57.1%. The model of Syngelaki et al. [20] had a specificity of 90.1 and 70.1% at a sensitivity of 23.7 and 68.4%, respectively.

Fig. 4. Decision curve analysis of the five best-performing models for preeclampsia. The solid grey line is the net benefit when considering all women as high risk and the horizontal black line when considering no women at high risk.

Fig. 3. Calibration plots of recalibrated first trimester prediction models for preeclampsia (PE). The grey line is the reference line with intercept = 0 and slope = 1 (perfect calibration). Triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF, correction factor.

Table 4. Predictive performance measures of the model of Syngelaki et al. [20] and effect of an intervention on preeclampsia incidence for different risk thresholds

Risk threshold ^a	High risk, % (n/n)	Sensitivity, % (n/n)	Specificity, % (n/n)	PPV, % (n/n)	NPV, % (n/n)	PE incidence reduction, % (NNP, n) Effect intervention ^b		
						RR 0.9	RR 0.8	RR 0.6
1.0%	76.2 (1,993/2,614)	98.7 (75/76)	24.4 (620/2,538)	3.8 (75/1,993)	99.8 (620/621)	9.9 (266)	19.7 (133)	39.5 (66)
2.0%	55.8 (1,459/2,614)	94.7 (72/76)	45.4 (1,151/2,538)	4.9 (72/1,459)	99.7 (1,151/1,155)	9.5 (203)	18.9 (101)	37.9 (51)
3.0%	37.3 (974/2,614)	76.3 (58/76)	63.9 (1,622/2,538)	6.0 (58/974)	98.9 (1,622/1,640)	7.6 (168)	15.3 (84)	30.5 (42)
4.0%	20.2 (528/2,614)	53.9 (41/76)	80.8 (2,051/2,538)	7.8 (41/528)	98.3 (2,051/2,086)	5.4 (129)	10.8 (64)	21.6 (32)
5.0%	12.4 (323/2,614)	38.2 (29/76)	88.4 (2,244/2,538)	9.0 (29/323)	97.9 (2,244/2,291)	3.8 (111)	7.6 (56)	15.3 (28)
6.0%	8.6 (224/2,614)	17.1 (13/76)	91.7 (2,327/2,538)	5.8 (13/224)	97.4 (2,327/2,390)	1.7 (172)	3.4 (86)	6.8 (43)

NNP, number needed to prevent; NPV, negative predictive value; PE, preeclampsia; PPV, positive predictive value; RR, risk reduction. ^a Predicted risk at or above this level was considered as high risk. ^b Intervention in all high-risk women.

Discussion

Main Findings

We assessed the generalizability of ten first trimester prediction models based on routinely collected maternal characteristics for the risk of PE in a Dutch population. Five models (Macdonald-Wallis et al. [17], Syngelaki et al. [20], Poon et al. [24], Poon et al. [25], Plasencia et al. [26]) showed moderate discriminative performance (AU-ROCs ranged from 0.73–0.77). The model of Poon et al. [25] showed the best calibration. This model underestimated the risk of developing PE over the entire range, but showed the best fit of predictor effects (slope 0.92). Refitting of the models improved calibration.

Interpretation

Validation is an important step in predictive modeling since the predictive performance is usually too optimistic in development samples [38]. Internal validation was only performed for three of the included models (Kenny et al. [18], North et al. [21] and Seed et al. [22]). The model of Kenny et al. [18] stayed most stable in our external validation cohort. External validation studies of prediction models for PE are scarce [13, 14, 39, 40]. None have simultaneously considered all first trimester prediction models based on routinely collected maternal characteristics for the risk of PE. Macdonald-Wallis et al. [17] estimated the external validity of their own developed model using another cohort from the same country. Independent external validation was only performed for the model of Plasencia et al. [26] and Poon et al. [25] [40–42]. These external validation studies, particularly those validating the model of Plasencia's group [41, 42], were hampered by a (very) low number of cases.

Our results confirm the importance of performing an external validation study. The predictive performance decreased for all included models. A previous pregnancy complicated by PE is a strong risk factor and the risk for multiparous women with no previous PE reduces substantially for a subsequent pregnancy [43–45]. The sub-analysis among nulliparous women showed that the performance decreased considerably, even for the models specifically developed for nulliparous women. This shows the difficulty of predicting PE among low risk nulliparous women based on exclusively routinely collected maternal characteristics. Although there appears scope for further improvement of the models, for now predictive performance of models based on only routinely collected maternal predictors measured up to 16 weeks of gestation is unlikely to improve sufficiently by adding other predictors as most studies already took into account the well-described maternal risk factors. Very strong risk factors such as antiphospholipid syndrome and chronic hypertension are absent in most models [45]. This is probably due to the low prevalence of some of these factors in the general pregnant population. Another possible factor is that pregnant women with essential hypertension may already receive high-risk treatment, reducing PE risk. Most prediction models are developed to assess the risk in a low-risk population. In case of the presence of very strong risk factors, women should also be considered as high risk.

Implications

Performance parameters are important characteristics of a prediction model but do not equal the usefulness of the model in clinical practice [38]. Moreover, a prediction model can only lead to benefit if there is an effective fol-

low-up strategy. Decision curve analysis indicated that the five best-performing models were clinically useful compared to treating all women or none at all as high risk over a small range of probability thresholds. No model had a considerably higher net benefit than others. Based on predictive performance, mainly among nulliparous women, we prefer the model of Syngelaki et al. [20] containing a wide range of known risk factors for PE (ethnicity, BMI, age, preexisting chronic hypertension, preexisting diabetes mellitus, smoking status, conception method, family history of PE, previous PE, and parity), this model also has good face value for the end user.

Decision curve analysis gives only a first impression of clinical usefulness. Determining the best risk threshold is a subsequent challenge that has had surprisingly little attention in the prediction literature. A prediction tool should correctly identify individuals at high risk of developing disease, while avoiding unnecessary interventions for individuals at low risk. There is a trade-off between sensitivity and specificity. In the case of PE, a high sensitivity is preferred due to the high burden of disease. The best performing models still had a considerably large number of false positives at a high sensitivity. The choice of a cut-off is therefore also dependent on the available intervention. Apart from the efficiency of the intervention in reducing PE, tolerance, side effects, and costs are as very important aspects to consider. The current available preventive interventions low-dose aspirin and calcium supplementation are safe, have no major side effects, are not burdensome and are relatively inexpensive [3–12].

Several current guidelines provide a list of risk factors to indicate women at high risk of developing PE. A limitation of some published guidelines is that they can be interpreted differently by caregivers [36, 46]. Furthermore, criteria lists do not take into account the strength of the different risk factors in relation to PE. Our best model performed better than the ACOG criteria and similarly to the NICE criteria at the same sensitivity.

At the moment, the validated models with an acceptable performance are simple, readily available, and cheap to implement. An extensive decision analysis can provide insight about the expected impact of the model compared to usual care by combining test characteristics with evidence on consequences of the outcome, effects and burden of the further management, and costs. In case the model is worth considering for use in clinical practice, an empirical impact study can determine whether the model indeed leads to better care. Not only outcomes are important but also a lot of other aspects like applicability of

the model, acceptability of the intervention, risk counseling, satisfaction, and real costs.

Strengths and Limitations

The main strengths of our study were the design and data collection. We performed a multicenter prospective study in which nearly all pregnant women less than 16 weeks of gestation were eligible. This design ensures a representative and unselected population to a high degree. The use of online questionnaires, an efficient data collection tool in women of reproductive age, improved our data quality and quantity of missing data due incorporation of validation checks [47]. Our dataset contained less than 1% missing data for almost all predictors, and these were imputed in order to prevent biased results and a loss of statistical precision. Outcome assessment based on both the postpartum questionnaires and medical records has ensured accurate and most complete available information possible in a large cohort [48].

We selected all existing first trimester prediction models based on routinely collected maternal characteristics for the risk of PE. The paper of Syngelaki et al. [20], included on the basis of citation lists, was not selected from our systematic search as the title and abstract did not clearly indicate that it addressed prediction models for a wide range of adverse pregnancy outcomes. Syngelaki's model was also not described in existing systematic reviews about prediction models for PE [13, 14, 39, 49]. A limitation in the selection process is that we had to exclude a substantial proportion of prediction models as the algorithm was not available, even after contacting the authors. Moreover, the dataset was too small to validate algorithms predicting early PE. Another point to be mentioned is that we used our own eligibility criteria for a fair comparison of all models. Analysis of the performance measures of the models according to their specific eligibility criteria however did not show better performance. Furthermore, we had to create a proxy variable for the predictors woman's own birth weight and crown-rump length. These proxy variables could have led to an underestimation of the predictive performance of the model. The variable blood pressure could have been subject to measurement error as this variable was self-reported and measured according to routine clinical practice [50]. Despite availability of standardized protocols for blood pressure measurement, the procedure is often not strictly followed in real practice. A last limitation to be mentioned is the number of cases in our validation cohort. No golden rule is available for the sample size of external validation studies. However, studies suggest that a minimum of

100 events, or preferably more, are needed for accurate and precise model performances [31, 51]. The study of Collins et al. [51] showed that the precision of the c-statistic and calibration slope did not differ considerably between 75 and 100 cases.

Conclusion

Five of the ten included first trimester prediction models based on routinely collected maternal characteristics for the risk of PE showed moderate predictive performance after external validation. The models are expected to perform better than current available guidelines in the separation of high- and low-risk pregnancies. Further research should focus on the determination of an acceptable risk threshold and impact of implementation of a chosen model in clinical practice.

Acknowledgements

We thank all women who participated in the Expect Study I. The Expect Study I could not have been established without the contribution of participating departments of obstetrics and gynecology of hospitals, midwifery practices, and maternity care centers in the Province of Limburg: Zuyderland Medical Center Heerlen and Sittard-Geleen, Maastricht University Medical Center, Laurentius Hospital Roermond, Sint Jans Gasthuis Weert, VieCuri

Medical Center, midwifery practice Roermond, midwifery practice Nederweert, midwifery practice Weert, midwifery practice Lennie & Chantal, midwifery practice Loes Wijnhoven, midwifery practice De Roerstreek, midwifery practice Bollebuik, midwifery practice Westenberg, midwifery practice Cranendonck, midwifery practice Becca, midwifery practice Born, midwifery practice Geleen, midwifery practice Grevenbicht, midwifery practice Sittard, midwifery practice Sittard-Oost, midwifery practice Stein, midwifery practice Astrea, midwifery practice Horst & Maasdorpen, midwifery practice Reuver-Tegelen, midwifery practice Janneke van Hal, midwifery practice Raijer & Sup, midwifery practice Venlo-Blerick, midwifery practice Venray, midwifery practice Schoffelen-Van Vleuten, midwifery practice Maastricht, midwifery practice Meerssen, midwifery practice Naomi Satijn, midwifery practice Vita, midwifery practice Het Verloskundig Huis, midwifery practice Valkenburg, midwifery practice Parkstad, midwifery practice Lief, midwifery practice Bevalt Beter, midwifery practice 't Bolleke, midwifery practice Natuurlijk bij Jeanny, midwifery practice La Vie, GroenekruisDomicura, Cicogna, ZiNkraamzorg, and maternity center Echt.

Disclosure Statement

The authors declare no conflicts of interest.

Funding Sources

The Expect Study was funded by The Netherlands Organization for Health Research and Development, Pregnancy and Childbirth program (ZonMw grant 209020007).

References

- Mol BW, Roberts CT, Thangaratinam S, Magee LA, de Groot CJ, Hofmeyr GJ: Pre-eclampsia. *Lancet* 2016;387:999–1011.
- Poon LC, Nicolaides KH: Early prediction of preeclampsia. *Obstet Gynecol Int* 2014;2014:297397.
- Hofmeyr GJ, Lawrie TA, Atallah AN, Duley L, Torloni MR: Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. *Cochrane Database Syst Rev* 2014;CD001059.
- Duley L, Henderson-Smith DJ, Meher S, King JF: Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane Database Syst Rev* 2007;CD004659.
- Askie LM, Duley L, Henderson-Smith DJ, Stewart LA: Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet* 2007;369:1791–1798.
- Bujold E, Roberge S, Lacasse Y, Bureau M, Audibert F, Marcoux S, Forest JC, Giguere Y: Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: a meta-analysis. *Obstet Gynecol* 2010;116:402–414.
- Roberge S, Nicolaides KH, Demers S, Villa P, Bujold E: Prevention of perinatal death and adverse perinatal outcome using low-dose aspirin: a meta-analysis. *Ultrasound Obstet Gynecol* 2013;41:491–499.
- Henderson JT, Whitlock EP, O'Connor E, Senger CA, Thompson JH, Rowland MG: Low-dose aspirin for prevention of morbidity and mortality from preeclampsia: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2014;160:695–703.
- Meher S, Duley L, Hunter K, Askie L: Antiplatelet therapy before or after 16 weeks' gestation for preventing preeclampsia: an individual participant data meta-analysis. *Am J Obstet Gynecol* 2017;216:121–128 e122.
- Roberge S, Nicolaides K, Demers S, Hyett J, Chaillet N, Bujold E: The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. *Am J Obstet Gynecol* 2017;216:110–120 e116.
- Roberge S, Bujold E, Nicolaides KH: Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. *Am J Obstet Gynecol* 2018;218:287–293 e1.
- Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, de Paco Matallana C, Akolekar R, Cicero S, Janga D, Singh M, Molina FS, Persico N, Jani JC, Plasencia W, Papaioannou G, Tenenbaum-Gavish K, Meiri H, Gizurarson S, Maclagan K, Nicolaides KH: Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. *N Engl J Med* 2017;377:613–622.
- Al-Rubaie Z, 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;123:1441–1452.

- 14 Henderson JT, Thompson JH, Burda BU, Cantor A: Preeclampsia screening: evidence report and systematic review for the us preventive services task force. *JAMA* 2017;317:1668–1683.
- 15 Steyerberg E: *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Berlin, Springer, 2008.
- 16 Justice AC, Covinsky KE, Berlin JA: Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–524.
- 17 Macdonald-Wallis C, Silverwood RJ, de Stavola BL, Inskip H, Cooper C, Godfrey KM, Crozier S, Fraser A, Nelson SM, Lawlor DA, Tilling K: 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:h5948.
- 18 Kenny LC, Black MA, Poston L, Taylor R, Myers JE, Baker PN, McCowan LM, Simpson NA, Dekker GA, Roberts CT, Rodems K, Noland B, Raymundo M, Walker JJ, North RA: Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. *Hypertension* 2014;64:644–652.
- 19 Direkvand-Moghadam A, Khosravi A, Sayehmiri K: Predictive factors for preeclampsia in pregnant women: a receiver operation character approach. *Arch Med Sci* 2013;9:684–689.
- 20 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;30:250–265.
- 21 North RA, McCowan LM, Dekker GA, Poston L, Chan EH, Stewart AW, Black MA, Taylor RS, Walker JJ, Baker PN, Kenny LC: Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ* 2011;342:d1875.
- 22 Seed PT, Chappell LC, Black MA, Poppe KK, Hwang YC, Kasabov N, McCowan L, Shenan AH, Wu SH, Poston L, North RA: 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;30:58–73.
- 23 Audibert F, Boucoiran I, An N, Aleksandrov N, Delvin E, Bujold E, Rey E: Screening for preeclampsia using first-trimester serum markers and uterine artery Doppler in nulliparous women. *Am J Obstet Gynecol* 2010;203:383.e381–e388.
- 24 Poon LC, 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;115:866–873.
- 25 Poon LC, Kametas NA, Pandeva I, Valencia C, Nicolaides KH: Mean arterial pressure at 11(+0) to 13(+6) weeks in the prediction of pre-eclampsia. *Hypertension* 2008;51:1027–1033.
- 26 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;30:742–749.
- 27 Meertens LJE, Scheepers HC, De Vries RG, Dirksen CD, Korstjens I, Mulder AL, Nieuwenhuijze MJ, Nijhuis JG, Spaanderman ME, Smits LJ: External validation study of first trimester obstetric prediction models (Expect Study I): research protocol and population characteristics. *JMIR Res Protoc* 2017;6:e203.
- 28 Robinson HP, Fleming JE: A critical evaluation of sonar "crown-rump length" measurements. *BJOG* 1975;82:702–710.
- 29 Brown MA, Hague WM, Higgins J, Lowe S, McCowan L, Oats J, Peek MJ, Rowan JA, Walters BN: The detection, investigation and management of hypertension in pregnancy: full consensus statement. *Aust N Z J Obstet Gynaecol* 2000;40:139–155.
- 30 Nederlandse Vereniging voor Obstetrie en Gynaecologie: Richtlijn Hypertensieve Aandoeningen in de Zwangerschap. Utrecht, Nederlandse Vereniging voor Obstetrie en Gynaecologie, 2005.
- 31 Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD: Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–483.
- 32 Van Buuren S: *Flexible Imputation of Missing Data*. Boca Raton, CRC Press, 2012.
- 33 Steyerberg EW, Vergouwe Y: Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–1931.
- 34 Vickers AJ, Elkin EB: Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–574.
- 35 National Collaborating Centre for Women's and Children's Health: *Hypertension in Pregnancy: The Management of Hypertensive Disorders during Pregnancy*. London, National Institute for Health and Clinical Excellence: Guidance, 2010.
- 36 LeFevre ML: Low-dose aspirin use for the prevention of morbidity and mortality from preeclampsia: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;161:819–826.
- 37 American College of Obstetricians and Gynecologists: *Practice Advisory on Low-Dose Aspirin and Prevention of Preeclampsia: Updated Recommendations (July 2016)*. Washington, American College of Obstetricians and Gynecologists, 2016.
- 38 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW: Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–138.
- 39 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.
- 40 Allen RE, Zamora J, Arroyo-Manzano D, Vela-luthar L, Allotey J, Thangaratinam S, Aquilina J: External validation of preexisting first trimester preeclampsia prediction models. *Eur J Obstet Gynecol Reprod Biol* 2017;217:119–125.
- 41 Farina A, Rapacchia G, Freni Sterrantino A, Pula G, Morano D, Rizzo N: Prospective evaluation of ultrasound and biochemical-based multivariable models for the prediction of late pre-eclampsia. *Prenat Diagn* 2011;31:1147–1152.
- 42 Herraiz I, Arbues J, Camano I, Gomez-Montes E, Graneras A, Galindo A: Application of a first-trimester prediction model for pre-eclampsia based on uterine arteries and maternal history in high-risk pregnancies. *Prenat Diagn* 2009;29:1123–1129.
- 43 Duckitt K, Harrington D: Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *BMJ* 2005;330:565.
- 44 Hernandez-Diaz S, Toh S, Cnattingius S: Risk of pre-eclampsia in first and subsequent pregnancies: prospective cohort study. *BMJ* 2009;338:b2255.
- 45 Bartsch E, Medcalf KE, Park AL, Ray JG: Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ* 2016;353:i1753.
- 46 World Health Organization: *WHO Recommendations for Prevention and Treatment of Pre-Eclampsia and Eclampsia*. Geneva, World Health Organization, 2011.
- 47 Van Gelder MM, Bretveld RW, Roeleveld N: Web-based questionnaires: the future in epidemiology? *Am J Epidemiol* 2010;kwq291.
- 48 van Gelder M, Vorstenbosch S, Derks L, Te Winkel B, van Puijenbroek EP, Roeleveld N: Web-based questionnaires to assess perinatal outcome proved to be valid. *J Clin Epidemiol* 2017;90:136–143.
- 49 Brunelli VB, Prefumo F: Quality of first trimester risk prediction models for pre-eclampsia: a systematic review. *BJOG* 2015;122:904–914.
- 50 Poon LC, Zymeri NA, Zamprakou A, Syngelaki A, Nicolaides KH: Protocol for measurement of mean arterial pressure at 11–13 weeks' gestation. *Fetal Diagn Ther* 2012;31:42–48.
- 51 Collins GS, Ogundimu EO, Altman DG: Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–226.