Pregnancy Hypertension 15 (2019) 42–50

Contents lists available at ScienceDirect

Pregnancy Hypertension

journal homepage: www.elsevier.com/locate/preghy

Temporal and external validation of the fullPIERS model for the prediction of adverse maternal outcomes in women with pre-eclampsia

U. Vivian Ukah[a,b,*], Beth Payne[b,c], Hanna Karjalainen[d], Eija Kortelainen[d], Paul T. Seed[e], Frances Inez Conti-Ramsden[e], Vivien Cao[f], Hannele Laivuori[d,g], Jennifer Hutcheon[a,b], Lucy Chappell[e], J. Mark Ansermino[b,c], Manu Vatish[h], Christopher Redman[h], Tang Lee[a], Larry Li[a], Laura A. Magee[e], Peter von Dadelszen[e], for the fullPIERS Group

[a] Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada
[b] BC Children's Hospital Research Institute, Vancouver, BC, Canada
[c] Department of Anaesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada
[d] Medical and Clinical Genetics, and Obstetrics and Gynaecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[e] School of Life Course Sciences, King's College London, London, UK
[f] Lower Mainland Pharmacy Services Residency Program, Vancouver, BC, Canada
[g] Department of Obstetrics and Gynecology, University of Tampere and Tampere University Hospital, Tampere, Finland
[h] Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK

ARTICLE INFO

Keywords:
Pregnancy hypertension
Pre-eclampsia
Prediction
Maternal outcomes
Prognosis
Model validation

ABSTRACT

The fullPIERS model is a risk prediction model developed to predict adverse maternal outcomes within 48 h for women admitted with pre-eclampsia. External validation of the model is required before implementation for clinical use. We assessed the temporal and external validity of the fullPIERS model in high income settings using five cohorts collected between 2003 and 2016, from tertiary hospitals in Canada, the United States of America, Finland and the United Kingdom. The cohorts were grouped into three datasets for assessing the primary external, and temporal validity, and broader transportability of the model. The predicted risks of developing an adverse maternal outcome were calculated using the model equation and model performance was evaluated based on discrimination, calibration, and stratification. Our study included a total of 2429 women, with an adverse maternal outcome rate of 6.7%, 6.6%, and 7.0% in the primary external, temporal, and combined (broader) validation cohorts, respectively. The model had good discrimination in all datasets: 0.81 (95%CI 0.75–0.86), 0.82 (95%CI 0.76–0.87), and 0.75 (95%CI 0.71–0.80) for the primary external, temporal, and broader validation datasets, respectively. Calibration was best for the temporal cohort but poor in the broader validation dataset. The likelihood ratios estimated to rule in adverse maternal outcomes were high at a cut-off of ≥30% in all datasets. The fullPIERS model is temporally and externally valid and will be useful in the management of women with pre-eclampsia in high income settings although model recalibration is required to improve performance, specifically in the broader healthcare settings.

1. Introduction

Pre-eclampsia and other hypertensive disorders of pregnancy (HDPs) remain a significant cause of maternal and fetal morbidity and mortality, globally [1,2]. Severe maternal morbidities resulting from pre-eclampsia include stroke, eclampsia and liver dysfunction [3]. Presently, delivery is the only cure; however, this is not always the best option for the fetus if the delivery occurs preterm [1]. While expectant management has been proposed as a means to achieve improved fetal survival, it is unclear for how long to delay delivery and how high the resultant risk is for the mother [4,5]. Accurate prognosis in women with pre-eclampsia and other HDPs is necessary to support a practice of expectant management and guide clinical decisions for timing of delivery, administration of antenatal corticosteroids and magnesium sulphate, or transfer to a higher level of care.

The fullPIERS (Pre-eclampsia Integrated Estimate of RiSk) model was developed to predict adverse maternal outcomes resulting from pre-eclampsia. The study's primary adverse maternal outcome was defined as one or more of the pre-specified severe maternal complications, which included central nervous system (CNS), hepatic, renal, cardiovascular

* Corresponding author at: 950 W 28th Avenue, Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC V5Z 4H4, Canada.
E-mail address: Vivian.Ukah@mail.mcgill.ca (U.V. Ukah).

https://doi.org/10.1016/j.preghy.2018.01.004
Received 7 September 2017; Received in revised form 21 December 2017; Accepted 4 January 2018
Available online 05 January 2018

and respiratory outcomes, occurring within 48 h of a woman's admission for pre-eclampsia (full definitions listed in Table S1) [6]. The rationale behind the PIERS project was that correctly identifying an individual woman's risk of complications before they happen would improve the clinician's ability to counsel that woman on timing of delivery and use of other interventions and avoid those complications. The multivariable model was developed in 2010 using a cohort of 2023 women admitted in tertiary centres in high income countries (HICs) with a 5% rate of adverse maternal outcomes. All the participating hospitals had expectant management policies for pre-eclampsia. Six predictor variables were included in the model: gestational age, chest pain or dyspnoea, oxygen saturation (SpO_2), platelet count, serum creatinine, and serum aspartate transaminase. The fullPIERS model was internally validated and had excellent discriminatory performance, with an area under the receiver-operating characteristic curve (AUROC) of 0.88 (95% confidence interval [CI], 0.84–0.92). The fullPIERS model equation is shown in Box 1. Although the model has promising predictive performance, temporal and external validation are necessary before it can be recommended for clinical use [7,8]. Validation ascertains the accuracy of the model in the population that it was designed for based on discriminatory and calibration performance. Temporal validation is carried out in the same setting as the one used during model development but with more recent patients, thereby prospectively evaluating model performance. External validation involves assessments in a similar population to that used for model development in other settings. Geographical validation is a type of external validation that includes cases that have the same inclusion and exclusion criteria as the development setting [8,9].

Although there have been a few attempts to externally validate the fullPIERS model, none of the studies conducted have used a similar patient populations and healthcare settings as the development population. Differences include use of data from low-and-middle-income countries (LMIC) [10,11] or using a subset (less than 34 weeks' gestational age) or broader inclusion criteria for disease (all HDPs). In addition, inadequate sample size has been an issue limiting the conclusions that can be drawn from these studies.[12] Using a sufficient sample size, we sought to assess geographic external- and temporal-validity of the fullPIERS model using datasets collected in high income countries with similar populations.

Box 1

The fullPIERS Logistic Regression Equation for the prediction of adverse maternal outcomes from pre-eclampsia: logit (pi) = 2·68 + (−5·41 × 10^{-2}; gestational age at eligibility) + 1·23(chest pain or dyspnoea) + (−2·71 × 10^{-2}; creatinine) + (2·07 × 10^{-1}; platelets) + (4·00 × 10^{-5}; platelets^2) + (1·01 × 10^{-2}; aspartate trans aminase) + (−3·05 × 10−6; AST^2) + (2·50 × 10^{-4}; creatinine × platelet) + (−6·99 × 10^{-5}; platelet × aspartate transaminase) + (−2·56 × 10^{-3}; platelet × SpO2)

2. Methods

2.1. Ethics

This study was approved by the University of British Columbia ethics board (CREB no: H07-02207).

2.2. Study population

2.2.1. Development cohort

The fullPIERS model was developed using data collected from high-income countries: Australia, Canada, New Zealand and the United Kingdom (UK). Data were collected prospectively from 2023 women admitted to a participating tertiary level facility with pre-eclampsia from July 2008 to January 2011. Pre-eclampsia was defined as hypertension with proteinuria or hyperuricaemia, or HELLP (Haemolysis Elevated Liver enzymes Low Platelet count) syndrome [6]. The study's primary adverse maternal outcome was a pre-specified composite of severe maternal complications occurring within 48 h of hospital admission. These outcomes were agreed upon using a Delphi consensus process prior to the model development study [6]. A full list of the outcomes is listed in Appendix S1 and a fullPIERS calculator online (https://pre-empt.cfri.ca/monitoring/fullpiers) and. Women were excluded from the cohort if they had already experienced an adverse maternal outcome before hospital admission and data collection or if they were admitted in spontaneous labour. Further details of the fullPIERS cohort have been described elsewhere [6].

2.2.2. Temporal and external validation datasets

We identified five cohorts for the temporal and external validation of the fullPIERS model. The decision to use these cohorts in our study was based on the cohorts having similar participant inclusion and exclusion criteria, and being derived in a similar tertiary unit health care setting. The data included were collected from the (i) British Columbia Women's (BCW) Hospital (ii) the Finnish Genetics of Preeclampsia Consortium (FINNPEC) study (iii) Pre-EcLampsIa: Clinical ApplicatioN (PELICAN) of PlGF study (iv) Alere-funded Pre-Eclampsia Triage by Rapid Assay (PETRA) study and, (v) John Radcliffe Hospital, Oxford (Oxford).

The BCW cohort was made up of data extracted from retrospective medical chart reviews for women admitted with pre-eclampsia to the BCW's Hospital (Vancouver, Canada) from January 2012 to May 2016. This site was one of the centres involved in the development of the fullPIERS model. The FINNPEC study recruited a cross-sectional, case–control cohort of singleton pregnancies from five university hospitals in Finland during 2008–2011.[13] The prospective PELICAN cohort included women who presented with symptoms or signs of suspected pre-eclampsia at one consultant-led maternity unit in the UK, January 2011 to February 2012 [14]. The BCW, FINNPEC and PELICAN cohorts centres practised expectant management for pre-eclampsia, similar to the centres involved in the development of the fullPIERS model. The PETRA cohort included women with symptoms and signs of pre-eclampsia presenting at twenty-four maternity units in the United States of America (USA) and Canada [15]. This was a prospective, observational cohort and women were recruited between January 2011 and February 2012; however, an interventionist management policy was more common in the Alere-PETRA centres. Finally, for the Oxford cohort, data were extracted retrospectively from the hospital flow sheets of women admitted in the Silver Star unit of John Radcliffe Hospital, Oxford, from 2003 to 2006.

For this study we used three sets of validation data: (i) the primary external data, comprising of the BCW, FINNPEC, and PELICAN cohorts, (ii) the temporal validation data, comprising of the BCW cohort only (iii) and the broader external data comprising of all five cohorts i.e., the BCW, FINNPEC, PELICAN, PETRA and Oxford cohorts. We divided the external validation into two sections – primary external validation using datasets with similar expectant management policies and data collection as that used during fullPIERS model development (to assess the reproducibility of the model in the most similar and reliable datasets) and "broader" external validation using combined datasets (to assess the overall transportability of the model) due to differences in clinical practice and data completeness. We planned *a priori* to base our interpretation of the model performance on the primary external dataset results. In addition, we assessed the model in only the BCW cohort for temporal validation because it was one of the model development sites.

2.3. Definition of pre-eclampsia and outcomes

We used the same definitions for pre-eclampsia as used in the fullPIERS model development study, as described above.

The primary outcome in our validation study was also the same as in the model development study (Table S1). When the exact time of day of the occurrence of an outcome was unknown, we used outcomes occurring within two calendar days from the date of admission as a proxy for adverse maternal outcomes occurring within 48 h. We examined outcomes occurring within seven days as a secondary outcome.

2.4. Statistical analyses

2.4.1. Demographics

The distribution of patient characteristics of the five cohorts included in the three validation sets were compared with the development (fullPIERS) cohort. Univariate comparisons of characteristics of women who experienced an adverse outcome and those who did not, were also performed for each validation data set.

2.5. Model performance evaluation

Using the worst value of each predictor measured within 48 h of admission prior to the occurrence of an adverse outcome, the published fullPIERS equation was applied to each of the validation datasets to calculate the predicted probability of experiencing an adverse outcome for each woman in the cohort under study. The performance of the model was evaluated based on its discrimination, calibration and stratification capacity.

Discriminative ability was assessed using the AUROC and was interpreted using the following pre-specified criteria: non-informative (AUROC ≤ 0·5), poor discrimination (0.5 < AUROC < 0·7), good discrimination (AUROC ≥ 0·7) [7,16]. Calibration was assessed by estimating the slope on a calibration plot of predicted versus observed outcome rates in each decile of predicted probability; a calibration slope of 1 and intercept of 0 is considered ideal [7]. Similar to the AUROC, a calibration slope was interpreted as: non-informative (slope ≤ 0·5), poor calibration (0.5 < slope < 0·7) and good calibration (slope ≥ 0·7) [17,18]. The stratification capacity of the model to classify the women into low- and high- risk categories was assessed using a classification table with generated risk groups (defined based on categories established in the model development study) [17,18]. The true and false positive rates, negative predictive values (NPVs), and positive predictive values (PPVs) were computed for each group. The Likelihood Ratios (LRs) were calculated for each group using the Deeks and Altman method for a multi-category diagnostic test [19].

2.5.1. Sensitivity analyses

For sensitivity analyses, we assessed the model performance in the validation cohorts for the prediction of adverse outcomes occurring within seven days of admission as done in the model development. This was done to evaluate the model's clinical utility within a longer time frame. Secondary analyses also included assessment of the discrimination capacity of the (i) primary and (ii) combined external validation data without the BCW's Hospital cohort. This was done to validate the inclusion of BCW data in all our analyses subsets.

2.5.2. Missing data

Missing SpO_2 was substituted with the median of 97%, as done during the fullPIERS model development [6]. If missing, AST was substituted with alanine transaminase (ALT) when available, as this measurement had been agreed to be biologically similar by expert opinion. For other variables or where both AST and ALT were absent, the type of missingness was explored by comparing cases with and without missing data in the validation cohorts, and multiple imputations were used to generate plausible values for missing variables. Multiple imputation was carried out ten times using the multiple imputation by chained equations (MICE) method [20–22].

2.5.3. Sample size

Our sample size was based on simulation studies which recommend 80–100 events (outcomes) and 100 non-events for sufficient power in validation studies [12]. This number of events was calculated to give 80% power at the 5% significance level. This was used to determine adequate statistical power in our study.

All statistical analyses were performed using R version 3·1·3 (The R Project for Statistical Computing).

3. Results

3.1. Comparison of the development and validation cohorts

In total, the combined cohort was made up of 2429 women: BCW (N = 1310), FINNPEC (N = 124), PELICAN (N = 70), PETRA (N = 644) and Oxford (N = 281). The distribution of patient characteristics between the development and individual validation cohorts are presented in Table 1. Compared to the development cohort, the women in the BCW cohort were more likely to be older, have a later onset of pre-eclampsia, and higher AST; the FINNPEC cohort was least likely to be multiparous and have more symptoms of chest pain or dyspnoea, while the PELICAN cohort had the lowest rate of smoking, and higher uric acid measurements. Compared to the development cohort, the women in the PETRA cohort were more likely to have an earlier onset of pre-eclampsia, had a higher rate of smoking, lower corticosteroid use for early onset pre-eclampsia but higher use of magnesium sulphate, and shorter admission to delivery for women with gestational age less than 34 weeks; they also had lower birth weights. The women in the Oxford cohort had higher platelets and creatinine measurements compared to the development cohort.

The combined distribution of patient characteristics for the cohorts grouped according to their analytical use (validation datasets), compared with the development cohort, are presented in Table S2. In total, the primary external, temporal validation and broader cohorts included 1504, 1310, and 2429 women respectively. Compared to the development cohort, the women in the primary external datasets were more likely to have an earlier onset of pre-eclampsia and less likely to smoke, while the women in the broader external datasets were more likely to multiparous and administered $MgSO_4$. The women in both the primary and broader external datasets were more likely to be older, compared to the development cohort.

Within 48 h of admission, the rates of adverse maternal outcomes encountered in the temporal, primary, and broader external validation cohorts were 87 (6.6%), 99 (6.7%) and 171 (7.0%), respectively. The rates of adverse maternal outcomes occurring within seven days or at any time during admission and the rates of stillbirths or neonatal deaths were similar between the validation and the development cohorts (Table S2).

Table S3 presents the individual components of the primary composite adverse outcome that occurred within 48 h of admission. The most common outcomes in combined validation cohorts were blood transfusion (N = 61), placental abruption (N = 21), and infusion of a third antihypertensive medication (N = 21). There were no cases of maternal deaths, cortical blindness or hepatic rupture.

3.2. Women with and without outcomes

In all the validation cohorts, women with an adverse maternal outcome within 48 h had an earlier onset of pre-eclampsia, worse clinical measures (i.e. higher chest pain, sBP, uric acid, and lower platelet count) and more interventions (antihypertensive and magnesium sulphate treatment) (Table S4). They also delivered at an earlier gestational age with babies of lower birth weights. These characteristics were similarly observed in the development cohort.[6]

Table 1
Maternal characteristics for the individual validation datasets and development dataset.

Characteristics	fullPIERS cohort (development) (2023 women)	BCW (1310 women) *(Also Temporal)*	FINNPEC (124 women)	PELICAN (70 women)	PETRA (644 women)	Oxford (2 8 1)
DEMOGRAPHICS & PREGNANCY CHARACTERISTICS						
Maternal age at EDD (yr)	31 [27,36]	34 [31,38]	31 [27, 34]	33 [29, 38]	30 [24, 34]	32 [28, 36]
Parity ≥ 1	581 (28·7%)	409 (31.2%)	25 (20.2%)	31 (44.3%)	280 (43.5%)	127 (45.2%)
Gestational age at eligibility (wk)**	36 [33, 38·3]	37.7 [35.6, 39]	35.2 [31.4, 37·0]	35.8 [34.3, 38.0]	33.9 [30.3, 36.3]	36.7 [33.7, 38.3]
Gestational age at eligibility < 34 weeks, N	636 (31.4%)	218 (16.6%)	51 (41.1%)	16 (22.9%)	331 (51.4%)	99 (35.2%)
Multiple pregnancy	192 (9·5%)	136 (10.4%)	0	5 (7.1%)	52 (8.1%)	26 (9.3%)
Smoking in this pregnancy	249 (12·3%)	90 (6.9%)	15 (12.1%)	2 (2.9%)	140 (21.7%)	23 (8.2%)
CLINICAL MEASURES						
Systolic BP (mm Hg)	160 [150, 176]	160 [151, 171]	169 [158, 179]	157 [150, 170]	143 [133, 154]	150 [140, 160]
Diastolic BP (mm Hg)	102 [98, 110]	100 [94, 105]	104 [99, 110]	98 [92, 102]	84 [76, 93]	98 [90, 101]
Chest pain/dyspnoea**	90 (4.4%)	82 (6.3%)	10 (8.1%)	3 (4.3%)	13 (2.0%)	7 (2.5%)
Uric acid (μM)	376 [320, 427]	379 [323, 436]	366 [323, 423]	400 [325, 495]	369 [309, 428]	337 [271, 390]
Lowest platelet count ($\times 10^9$ per L)**	192 [150, 242]	174 [136, 217]	187 [153, 232]	170 [128, 212]	203 [158, 248]	225 [178, 275]
Highest AST/ALT (U/L)**	28 [21, 41]	33 [26, 47]	19 [14, 30]	20 [14, 32]	23 [18, 34]	17 [13, 27]
Creatinine (μM)	67 [58, 77]	64 [56, 75]	62 [54, 69]	70 [59, 84]	61 [53, 71]	75 [68, 84]
INTERVENTIONS DURING ADMISSION						
Corticosteroids	550 (27·2%)	320 (24.4%)	56 (45.2%)	31 (44.3%)	161 (25.0%)	78 (27.8%)
Corticosteroids, GA onset < 34	440/636 (69.2%)	195/218 (89.5%)	43/51 (84.3%)	14/16 (87.5%)	137/331 (41.4%)	65/99 (65.7%)
Antihypertensive therapy	1381 (68·3%)	896 (68.4%)	104 (83.9%)	58 (82.9%)	463 (71.9%)	175 (62.3%)
$MgSO_4$	690 (34·1%)	393 (30.0%)	69 (55.7%)	11 (15.7%)	464 (72.1%)	31 (11.0%)
PREGNANCY OUTCOMES						
Admission-To-Delivery Interval (Days)	2 [1, 5]	1 [1, 3]	4 [2,7]	6 [3, 14]	2 [1, 4]	3 [1, 8]
Admission-To-Delivery Interval, $< 34^{+0}$ Weeks (Days)	4 [2, 14]	4 [2, 11]	6 [3, 8]	13 [8, 26]	3 [1, 6]	8 [4, 19]
Gestational age at delivery (wk)	36.9 [34·1, 38·6]	37.8 [36, 39.1]	35.9 [32.3, 37.9]	37.6 [36.3, 38.3]	34.6 [31.1, 36.9]	36.7 [33.7, 38.3]
Birth weight (grams)	2141 [1441, 2807]	2885 [2275, 3364]	2305 [1475, 2930]	2700 [2065, 3150]	2070 [1286, 2770]	2516 [1647, 3216]
Stillbirth	20 (1.0%)	7 (0.5%)	0	0	10 (1.6%)	8 (2.9%)
Neonatal death	26 (1·3%)	10 (0.8%)	1 (0.8%)	0	13 (2.0%)	25 (8.9%)
MATERNAL OUTCOME(N women)						
Within 48 h	106 (5.2%)	87 (6.6%)	11 (8.9%)	1 (1.4%)	48 (7.5%)	24 (8.5%)
Within 7 days	203 (10.0%)	110 (8.4%)	40 (32.3%)	2 (2.9%)	56 (8.7%)	45 (16.0%)
At anytime	261 (12.9%)	122 (9.3%)	62 (50.0%)	6 (8.6%)	62 (9.6%)	57 (20.3%)

AST (aspartate aminotransferase), BP (blood pressure), EDD (estimated date of delivery), $MgSO_4$ (magnesium sulphate).
**Variables included in the model.

3.3. Data completeness

Table S5 shows the number of missing predictor variables in each of the analysis/validation cohorts. Gestational age at admission for disease was the most complete variable in all the validation datasets except for two missing cases (0.1%) in the broader external dataset while chest pain/dyspnoea and SpO_2 had the highest proportion of missing data. The broader validation dataset had the most missingness overall with 3.3% for platelet count, 4.5% for AST or ALT, 7.3% for serum creatinine, 37.2% for chest pain or dyspnoea, and 42.4% for SpO_2.

3.4. Model performance

3.4.1. Primary external validation

The fullPIERS model showed good discrimination in the primary external validation dataset with an AUROC of 0.81 (95% CI 0.76–0.87) (Fig. 1a). Imputation of missing variables did not show any significant change in the discriminatory performance (AUROC of 0.81 (95% CI 0.75–0.86).

The model also showed good calibration performance in the primary external validation dataset with a slope of 0.70, although the intercept was marginally elevated ($\alpha = 0.3$) (Fig. 2a).

Table 2 shows the distribution of women in pre-specified predicted probability risk groups and the stratification capacity of the model in each group for the primary external validation dataset. The distribution of women in these risk group was similar to the model development study with 30.5% of women having a predicted probability of < 1% (35% in the development cohort) and 4% of women with a predicted probability of ≥30% (also 4% in the development cohort) [6]. Using the predicted probability cut-off of ≥30% for high risk (pre-identified threshold in the model development study), 55% of the women had an adverse outcome. The resulting false positive rate was 2% (specificity of 98%) and the true positive rate (sensitivity) was 36% with a high LR of 17 (95%CI 10.97–26.43) showing strong evidence to rule in adverse maternal outcomes; the LR at the lower predicted scores (< 2.5%) were not useful for ruling out adverse outcomes. Overall, the model was able to stratify women into a high risk group (predicted probability ≥30%) and a low risk group (predicted probability < 30%).

3.4.2. Temporal validation

For the temporal validation using only the BCW cohort, the fullPIERS model showed good discrimination capacity with an AUROC of 0.82 (95% CI 0.76–0.87) (Fig. 1c), which did not change after imputation of missing values. Calibration was good with a slope of 0.70 and intercept of 0.20 (Fig. 2b).

Table 3 presents the distribution of women and the stratification capacity of the model in the temporal validation datasets in each pre-specified predicted risk group. The proportions of women in the lowest (< 1%) and highest risk group (≥30%) were 30% and 5% respectively. The proportion of women with adverse outcomes in the highest risk group of ≥ 0.3 (56%) was also similar to the proportion in the development data (59%).[6] The resulting false positive rate was 2% (specificity 98%) and the true positive rate was 41% with a high LR of 18 (95%CI 11.60–28.16) at this threshold, showing strong evidence to rule in adverse maternal outcomes.

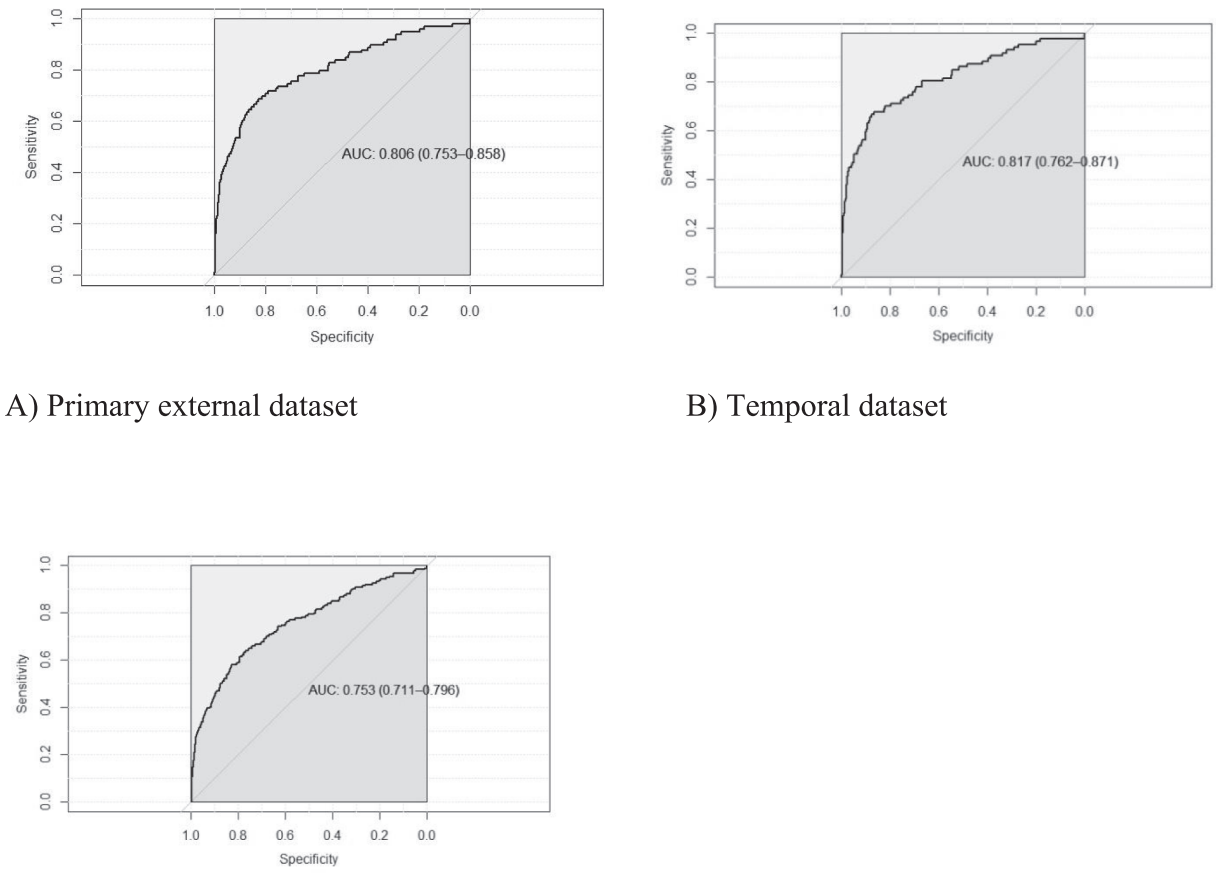

Fig. 1. Discrimination performance of the fullPIERS model within 48 h of admission.

3.4.3. Broader external validation

The fullPIERS model retained good discrimination in the broader validation dataset although the AUROC decreased (0.74 (95% CI 0.69–0.80) (Fig. 1c). As seen in the prior validation cohorts, there was also no significant change after imputation of missing data (AUROC of 0.75 (95% CI 0.71–0.80). Calibration ability was poor with a slope of 0.55 and intercept of 0.30 (Fig. 2c).

Similar to the primary external validation dataset, about 4% of women had a predicted probability of ≥30% (Table 4) [6]. In this highest risk group, half of women had an adverse outcome. The resulting false positive rate was 2% (specificity 98%) and the true positive rate was 27% with a LR of 13 (95% CI 9.21–18.9).

Thus, the LR also showed strong evidence to rule in adverse maternal outcomes in all datasets.

3.5. Sensitivity analyses

3.5.1. Performance for outcomes within seven days

Model performance decreased in all three validation datasets for the prediction of adverse outcomes within seven days. The AUROCs after imputation of missing data were 0.71 (95% CI 0.66–0.76), 0.69 (95% CI 0.65–0.73), and 0.78 (95% CI 0.72–0.83), for the primary external, broader and temporal validation datasets, respectively. Calibration was poor in all datasets with slopes of 0.63, 0.61 and 0.44 for the temporal, primary external and broader external datasets, respectively. Similar results were observed from the complete case analyses.

3.5.2. Performance of model excluding the BCW cohort

The discriminatory performance of the model dropped in both the primary and broader external validation datasets upon the exclusion of the BCW's cohort, although the AUROCs remained > 0.70 (Appendix Fig. 1).

4. Discussion

We assessed the temporal and external validation of the fullPIERS model using three datasets from high income countries for the prediction of adverse maternal outcomes within 48 h. Overall, the model retained good discriminatory performance across all datasets (AUROC ≥ 0.7). Calibration was good in the primary and temporal external datasets but was poor in the broader external dataset, as reflected by the reduction in slope. An increase in the calibration intercept was also observed in all three of the validation cohorts. Despite errors in calibration the model was able to classify women into low- and high-risk groups using a predicted probability cut-off of ≥30% and showed a strong ability to 'rule in' adverse outcomes within 48 h at this cut-off.

In external validation studies, decreases in model performance are common and can be a result of overfitting of the model to the data used for development, case-mix differences between the development and validation cohorts, or differences in the effect of the model predictors between the development and validation cohorts, or a combination of some or all of these factors [8]. All the AUROCs estimated in the three validation cohorts were lower than that estimated in the original fullPIERS model development study (0.88 (95% CI 0.84–0.92). This difference was only significant in the broader validation cohort (the confidence intervals overlapped for the primary external and temporal

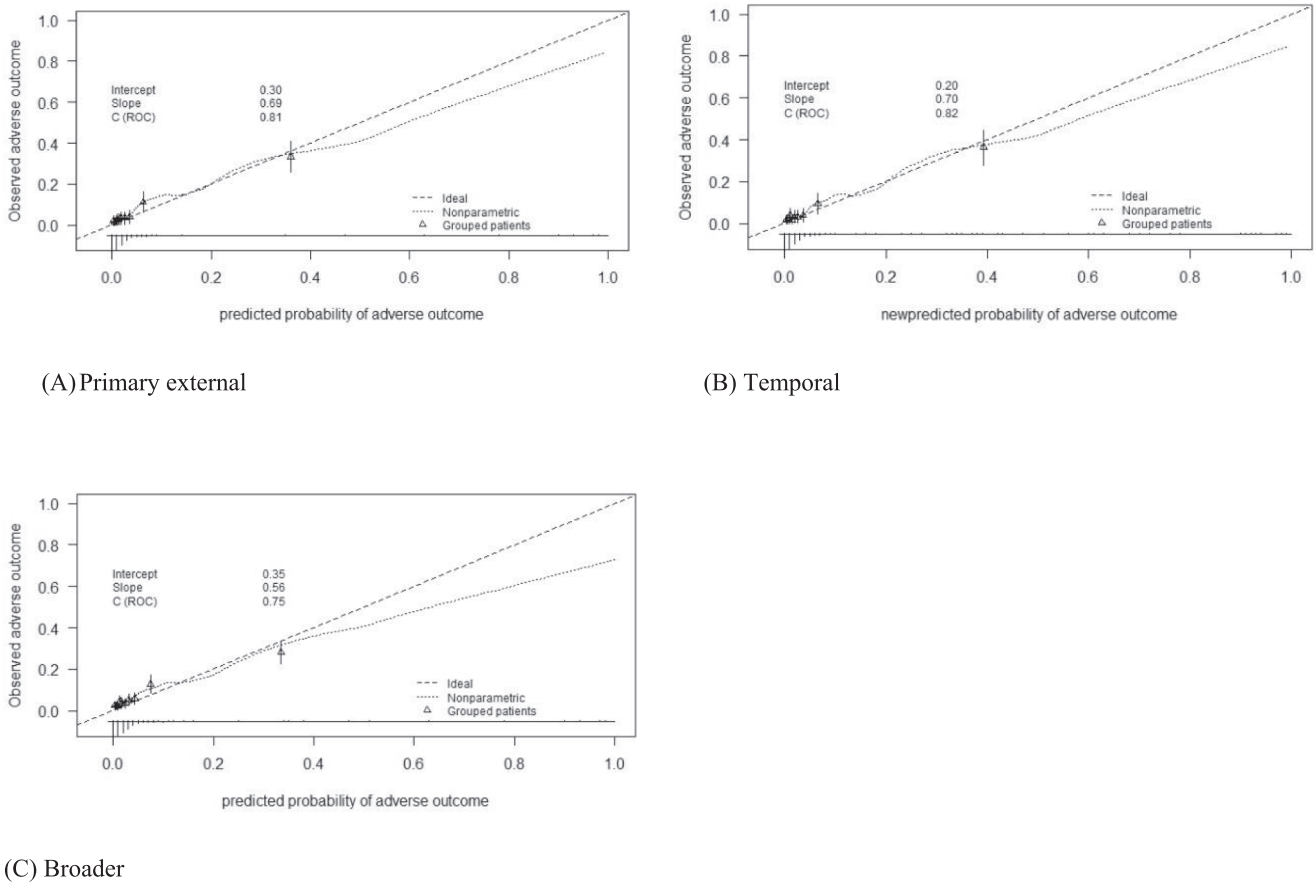

Fig. 2. Calibration performance of the fullPIERS model within 48 h of admission.

validation). A decrease in performance was also observed in the calibration ability of the model upon temporal validation, whereby the dataset most similar to the development cohort was used.

There were fewer case-mix differences between the development and temporal validation data with regard to demographics, antihypertensive administration, and adverse neonatal and maternal outcomes (Table 1). Despite the fact that the women in the temporal cohort had worse predictor measurements, the rates of adverse outcomes between the two cohorts were similar, suggesting a possible difference in the predictor-outcome relationship. Studies have reported that a slope < 1 is indicative of inconsistent predictor effects or/and overfitting [7,23]. We suspect that it is more likely that the reduced performance was due to different predictor effects and perhaps, overfitting than due to case-mix differences [9,23].

In addition to the predictor-effects, case-mix differences may have played a more substantial role in the broader external dataset, as there were more differences in population characteristics observed in the PETRA and Oxford datasets compared to the fulPIERS cohort. There was also a slightly higher rate of adverse maternal outcome within 48 h, which was reflected in the calibration intercept (0.3). Of note, the PETRA cohort, which significantly contributed to the broader validation dataset, primarily included women admitted in the USA. The

Table 2
Risk stratification table to assess the performance of the fullPIERS model for predicting maternal outcome at varying predicted probability cut-off values within 48 h in the Primary External Validation dataset.

Prediction score range (Mean predicted probability)	Total N women in range (%) (N = 1504)	N women with outcome (% (N = 99)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	*True positive rate (%) [95% CI]	False positive rate (%) [95% CI]
< 1.0% (0.5%)	459 (30.5%)	8 (1.7%)	0·25 [0·12–0.52]	–	–	–	–
1·0–2·4% (1.2%)	498 (33.1%)	14 (2.8%)	0·41 [0·25–0.67]	98 [0·96–0·99]	8.6 [0·07–0·11]	92 [0·84– 0·96]	68 [0·65–0·70]
2·5–4·9% (3.4%)	287 (19.1%)	13 (4.5%)	0·67 [0·40–1.13]	98 [0·96–0·99]	14 [0·11–0·17]	78 [0·68–0·85]	34 [0·31– 0·40]
5·0–9·9% (7.0%)	117 (7.8%)	16 (13.7%)	2.25 [1.38–3.66]	97 [0·96–0·98]	25 [0·20–0·31]	65 [0·54–0·74]	14 [0·12–0·16]
10·0–29·9% (17.2%)	77 (5.1%)	12 (15.6%)	2.62 [1.47–4.68]	96 [0·95–0·97]	47 [0·36–0·58]	41 [0·32–0·52]	7 [0·06–0·08]
≥ 30·0% (60.9%)	66 (4.4%)	36 (54.5%)	17.03 [10.97–26.43]	96 [0·94–0·97]	55 [0·42–0·67]	36 [0·27–0·47]	2 [0·02–0·03]

LR, Likelihood ratios; PPV, positive predictive value; NPV, negative predictive value.
*True positive rate (or Sensitivity), false positive rate (1-Specificity).

Table 3
Risk stratification table to assess the performance of the fullPIERS model for predicting maternal outcome at varying predicted probability cut-off values within 48 h in the temporal dataset.

Prediction score range	Total N women in range (%) (N = 1310)	N women with outcome (%) (N = 87)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	*True positive rate (%) [95% CI]	False positive rate (%) [95% CI]
< 1.0% (0.6%)	395 (30.2%)	6 (1.5%)	0·22 [0·09–0·50]	–	–	–	–
1·0–2·4% (1.2%)	430 (32.8%)	11 (2.6%)	0·37 [0·21–0·64]	98 [0·97–0·99]	8.9 [0·07–0·11]	93 [0·85– 0·97]	68 [0·66–0·70]
2·5–4·9% (3.4%)	247 (18.9%)	11 (4.5%)	0·66 [0·37–1.51]	98 [0·97–0·99]	14 [0·12–0·18]	80 [0·70–0·88]	34 [0·31– 0·37]
5·0–9·9% (7.0%)	109 (8.3%)	14 (12.8%)	2.07 [1.24–3.47]	97 [0·96–0·98]	25 [0·20–0·31]	68 [0·57–0·77]	15 [0·13–0·17]
10·0–29·9% (16.9%)	65 (5.0%)	9 (13.8%)	2.26 [1.16–4.41]	96 [0·95–0·97]	48 [0·37–0·59]	45 [0·34–0·56]	7 [0·06–0·08]
≥ 30·0% (61.5%)	64 (4.9%)	36 (56.3%)	18.07 [11.60–28·16]	96 [0·95–0·97]	56 [0·43–0·68]	41 [0·31–0·52]	2 [0·02–0·03]

LR, Likelihood ratios; PPV, positive predictive value; NPV, negative predictive value.
*True positive rate (or Sensitivity), false positive rate (1-Specificity).

Table 4
Risk stratification table to assess the performance of the fullPIERS model for predicting maternal outcome at varying predicted probability cut-off values within 48 h in the Broader (Combined) dataset.

Prediction score range	Total N women in range (%) (N = 2429)	N women with outcome (%) (N = 171)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	*True positive rate (%) [95% CI]	False positive rate (%) [95% CI]
< 1% (0.6%)	604 (24.9%)	14 (2.3%)	0·31 [0·18–0·55]	–	–	–	–
1·0–2·4% (1.7%)	779 (32.1%)	27 (3.5%)	0·47 [0·34–0·66]	98 [0·96–0·99]	8.6 [0·07–0·10]	92 [0·86– 0·95]	74 [0·72–0·76]
2·5–4·9% (3.5%)	526 (21.7%)	31 (5.9%)	0·83 [0·61–1.12]	97 [0·96–0·99]	12 [0·11–0·15]	76 [0·69–0·82]	41 [0·39– 0·43]
5·0–9·9% (7.0%)	259 (10.7%)	30 (11.6%)	1.73 [1.25–2.40]	96 [0·95–0·97]	19 [0·16–0·23]	58 [0·50–0·65]	19 [0·17–0·20]
10·0–29·9% (16.3%)	169 (7.0%)	23 (13.6%)	2.08 [1.40–2.09]	95 [0·94–0·96]	26 [0·21–0·32]	40 [0·33–0·48]	9 [0·07–0·10]
≥ 30·0% (60.4%)	92 (3.8%)	46 (50.0%)	13.20 [9.21–18.9]	95 [0·94–0·96]	50 [0·39–0·61]	27 [0·21–0·34	2 [0·02–0·03]

LR, Likelihood ratios; PPV, positive predictive value; NPV, negative predictive value.
*True positive rate (or Sensitivity), false positive rate (1-Specificity).

pattern of practice for the management of women with pre-eclampsia in the USA is different from the other datasets in that it is more interventionist than expectant; this could also be observed from the short admission to delivery interval for women with gestational age less than 34 weeks compared to the other cohorts (Table 1). Earlier delivery would shorten the natural course of pre-eclampsia and could reduce performance of the model. The extreme predictions shown in the calibration graph (over-prediction of outcomes in the lower risk groups and under-prediction in the higher risk groups) as well as the significant reduction in slope (0.55) are also suggestive of overfitting of the model [9]. These findings suggest a need for recalibration of the model to improve its performance in broader external dataset.

4.1. Comparison with existing literature

Four previous studies have assessed the validity of the fullPIERS model [10,11,24]. The study by Akkermans et al. [24]. used a cohort of women with severe, early-onset pre-eclampsia admitted into tertiary centres in the Netherlands. Although this study included patients from a similar setting to the fullPIERS cohort i.e. tertiary, high income setting, their inclusion criteria may have resulted in a significant case-mix difference. In contrast with our study, they reported a higher discriminative performance of the model (AUC ROC 0.97, 95% CI: 0.94–0.99). The validation studies by Agrawal et al. [10] and Ukah et al. [11] both examined the fullPIERS model in low resourced settings and the study by Hadley et al. [25] included women with all HDPs. The use of datasets with populations from low-and-middle income settings (less resources and higher rates of outcomes), or with different inclusion criteria (e.g. other types of HDPs were included), and different management practice (e.g. less expectant management), compared to the development cohort, may have also contributed to an increase in severity of case-mix and resulted in the lower performance reported by these studies. Similar to our study; these three studies [10,11,25] reported a decrease in discriminative ability (although all three studies still reported good discrimination AUROC > 0.7).

4.2. Strengths and limitations

A strength of this study is the assessment of the validity and transportability of the fullPIERS risk prediction model using data similar to the model development cohort i.e. data from tertiary and high-income settings. We also had sufficient power to detect any major changes in the model performance. The findings from our analyses are important to actually determine if the model itself is valid as developed, and not just if it works by chance or due to peculiarities in the validation cohort used. The combination of cohorts from different sites also makes our findings more generalizable; thus our findings represent a true validation of the model's performance in similar settings.

One limitation is that to achieve adequate power, we included women from BCW's Hospital, which was one of the development sites, in the primary and broader external validation datasets. Although it may be ideal to use completely different sites for an external validation, these women were enrolled in a later time period from those used in the development study and can still be considered an external cohort. Our sensitivity analyses assessing the impact of including the BCW cohort demonstrated that the discriminatory performance without the BCW cohort remained good (AUROC > 0.70) showing that the model performance was not entirely dependent on the BCW cohort.

Another limitation was the large percentage of missing data in the broader external validation dataset. Although we accounted for this limitation by using multiple imputation techniques, these may have affected the precision of the model performance estimate. However, research suggests that imputation is preferable to omission of individuals, even if a predictor is completely missing in a dataset [22,26]. Our imputation analyses did not show any significant difference, thereby suggesting that there was less likelihood of bias in the broader external datasets. We hope that our results will encourage the measurements of SpO_2 and other model variables since they are important predictors of adverse maternal outcomes within 48 h.

4.3. Implications for clinical practice

This external validation study shows that the fullPIERS model is useful in discriminating between patients at high and low risk of adverse maternal outcomes within 48 h and even up to a week after assessment. Our study also shows that using a threshold of ≥30% predicted probability was a good threshold to rule-in the outcome. Based on our results, the model can be used to aid clinicians in managing women with pre-eclampsia in similar settings and to make decisions such as transfer to higher care units and delivery. However, caution should be applied when using the model in settings with a broader case-mix of patients or a more interventionist management style such as those participating in the PETRA study. Recalibration of the model should be considered in these settings before clinical use.

5. Conclusion

The fullPIERS model is temporally and externally valid for the prediction of adverse maternal outcomes occurring within 48 h of admission for pre-eclampsia. Recalibration might be helpful in improving the calibration performance in more diverse settings. Future studies should focus on recalibration and assessing the model performance in broader sub-groups.

Declaration of interests

None.

Details of ethics approval

This study was approved by the Clinical Research Ethics Board at the University of British Columbia on March 1, 2014 (UBC CREB number: H07-02207).

Funding

This study was supported by the Canadian Institutes of Health Research (CIHR operating grants). The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Acknowledgements

We are grateful to all who have contributed to data collection, expertise and collaboration in this study: Dr. Paul Sheard, Sharla Drebit, Domena Tu, Lillian Cao, Katherine Thomas, Phenicia Azurin, Matthew Haslam, Benita Okocha, Layla Lavallee, Elizabeth Wilcox, Ryan Ingram, Sheryl Atkinson, Dr. Doug Woelkers, Dr. Kenneth Kupfer, Dr. Baha Sibai and the Alere-PETRA group.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.preghy.2018.01.004.

References

[1] P. von Dadelszen, L.A. Magee, Pre-eclampsia: an update, Curr. Hypertens Rep. 16 (8) (2014) 1–14, https://doi.org/10.1007/s11906-014-0454-8.
[2] P. von Dadelszen, L.A. Magee, Preventing deaths due to the hypertensive disorders of pregnancy, Best Pract. Res. Clin. Obstetrics Gynaecol. 36 (2016) 83–102, https://doi.org/10.1016/j.bpobgyn.2016.05.005.
[3] J.A. Hutcheon, S. Lisonkova, K.S. Joseph, Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy, Best Pract. Res. Clin. Obstetrics Gynaecol. 25 (4) (2011) 391–403, https://doi.org/10.1016/j.bpobgyn.2011.01.006.
[4] K. Broekhuijsen, G.J.v. Baaren, M.G.v. Pampus, et al., Immediate delivery versus expectant monitoring for hypertensive disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an open-label, randomised controlled trial, Lancet 385 (9986) (2015) 2492–2501, https://doi.org/10.1016/S0140-6736(14)61998-X.
[5] C.M. Koopmans, D. Bijlenga, H. Groen, et al., Induction of labour versus expectant monitoring for gestational hypertension or mild pre-eclampsia after 36 weeks' gestation (HYPITAT): a multicentre, open-label randomised controlled trial, Lancet 374 (9694) (2009) 979–988, https://doi.org/10.1016/S0140-6736(09)60736-4.
[6] P. von Dadelszen, B. Payne, J. Li, et al., Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model, Lancet 377 (9761) (2011) 219–227, https://doi.org/10.1016/S0140-6736(10)61351-7.
[7] E.W. Steyerberg, A.J. Vickers, N.R. Cook, et al., Assessing the performance of prediction models: A framework for traditional and novel measures, Epidemiology 21 (1) (2010) 128–138, https://doi.org/10.1097/EDE.0b013e3181c30fb2.
[8] T. Neeman, Clinical prediction models: A practical approach to development, validation, and updating by ewout W. steyerberg, Int. Stat. Rev. 77 (2) (2009) 320–321.
[9] Y. Vergouwe, C. Moons, E. Steyerberg, External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients, Am. J. Epidemiol. 172 (8) (2010) 971–980, https://doi.org/10.1093/aje/kwq223.
[10] S. Agrawal, N. Maitra, Prediction of adverse maternal outcomes in preeclampsia using a risk prediction model, J. Obstet. Gynaecol. India 66 (Suppl 1) (2016) 104.
[11] U.V. Ukah, B. Payne, T. Lee, L.A. Magee, P. von Dadelszen, External validation of the fullPIERS model for predicting adverse maternal outcomes in pregnancy hypertension in low- and middle-income countries, Hypertension (2017) HYPERTENSIONAHA.116.08706. doi: 10.1161/HYPERTENSIONAHA.116.08706.
[12] Y. Vergouwe, E.W. Steyerberg, M.J.C. Eijkemans, J.D.F. Habbema, Substantial effective sample sizes were required for external validation studies of predictive logistic regression models, J. Clin. Epidemiol. 58 (5) (2005) 475–483, https://doi.org/10.1016/j.jclinepi.2004.06.017.
[13] T. Jääskeläinen, S. Heinonen, E. Kajantie, J. Kere, K. Kivinen, A. Pouta, H. Laivuori, FINNPEC Study Group. Cohort profile: the Finnish Genetics of Pre-eclampsia Consortium (FINNPEC), BMJ Open 6 (2016) e013148, https://doi.org/10.1136/bmjopen-2016-013148.
[14] L.C. Chappell, S. Duckworth, P.T. Seed, et al., Diagnostic accuracy of placental growth factor in women with suspected preeclampsia: a prospective multicenter study, Circulation 128 (19) (2013) 2121–2131, https://doi.org/10.1161/CIRCULATIONAHA.113.003215.
[15] D.A. Woelkers, P. von Dadelszen, B. Sibai, 482: Diagnostic and prognostic performance of placenta growth factor (PLGF) in women with signs or symptoms of early preterm preeclampsia, Obstet. Gynecol. 214 (1) (2016) S264, https://doi.org/10.1016/j.ajog.2015.10.525.
[16] K. Van Hoorde, Y. Vergouwe, D. Timmerman, S. Van Huffel, E.W. Steyerberg, B. Van Calster, Assessing calibration of multinomial risk prediction models, Stat. Med. 33 (15) (2014) 2585–2596, https://doi.org/10.1002/sim.6114.
[17] M.S. Pepe, Z. Feng, H. Janes, P.M. Bossuyt, J.D. Potter, Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design, J. Natl. Cancer Inst. 100 (20) (2008) 1432–1438, https://doi.org/10.1093/jnci/djn326.
[18] H. Janes, M.S. Pepe, W. Gu, Assessing the value of risk predictions by using risk stratification tables, Ann. Int. Med. 149 (10) (2008) 751.
[19] J.J. Deeks, D.G. Altman, Diagnostic tests 4: likelihood ratios, BMJ 329 (7458) (2004) 168–169, https://doi.org/10.1136/bmj.329.7458.168.
[20] Buuren V. Stef, K. Groothuis-Oudshoorn, Mice: multivariate imputation by chained equations in R, J. Stat. Softw. 45 (3) (2011).
[21] R.C.A. Rippe, M. den Heijer, S. le Cessie, Imputation of missing data, Ned. Tijdschr.

Geneeskd. 157 (18) (2013) A5539.
[22] P. Cummings, Missing data and multiple imputation, JAMA Pediatrics 167 (7) (2013) 656–661, https://doi.org/10.1001/jamapediatrics.2013.1329.
[23] T.P. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E.W. Steyerberg, K.G.M. Moons, A new framework to enhance the interpretation of external validation studies of clinical prediction models, J. Clin. Epidemiol. 68 (3) (2015) 279–289, https://doi.org/10.1016/j.jclinepi.2014.06.018.
[24] J. Akkermans, B. Payne, P. von Dadelszen, et al., Predicting complications in pre-eclampsia: External validation of the fullPIERS model using the PETRA trial dataset, Eur. J. Obstet. Gynecol. Reproductive Biol. 179 (2014) 58–62, https://doi.org/10.1016/j.ejogrb.2014.05.021.
[25] E.E. Hadley, A. Poole, S.R. Herrera, et al., 472: External validation of the fullPIERS (preeclampsia integrated estimate of RiSk) model, Obstet. Gynecol. 214 (1) (2016), https://doi.org/10.1016/j.ajog.2015.10.515 S259-S260-S260.
[26] U. Held, A. Kessels, J. Garcia Aymerich, et al., Methods for handling missing variables in risk prediction models, Am. J. Epidemiol. 184 (7) (2016) 545–551, https://doi.org/10.1093/aje/kwv346.