HYPERTENSION IN PREGNANCY

Hypertension in Pregnancy

VOLUME 38
NUMBER 2
2019

ISSN: 1064-1955 (Print) 1525-6065 (Online) Journal homepage: https://www.tandfonline.com/loi/ihip20

External validation of prognostic models for preeclampsia in a Dutch multicenter prospective cohort

Marije Lamain-de Ruiter, Anneke Kwee, Christiana A. Naaktgeboren, Rebecca D. Louhanepessy, Inge De Groot, Inge M. Evers, Floris Groenendaal, Yolanda R. Hering, Anjoke J. M. Huisjes, Cornel Kirpestein, Wilma M. Monincx, Peter C. J. I. Schielen, Annewil Van 'T Zelfde, Charlotte M. Van Oirschot, Simone A. Vankan-Buitelaar, Mariska A. A. W. Vonk, Therese A. Wiegers, Joost J. Zwart, Karel G. M. Moons, Arie Franx & Maria P. H. Koster

To cite this article: Marije Lamain-de Ruiter, Anneke Kwee, Christiana A. Naaktgeboren, Rebecca D. Louhanepessy, Inge De Groot, Inge M. Evers, Floris Groenendaal, Yolanda R. Hering, Anjoke J. M. Huisjes, Cornel Kirpestein, Wilma M. Monincx, Peter C. J. I. Schielen, Annewil Van 'T Zelfde, Charlotte M. Van Oirschot, Simone A. Vankan-Buitelaar, Mariska A. A. W. Vonk, Therese A. Wiegers, Joost J. Zwart, Karel G. M. Moons, Arie Franx & Maria P. H. Koster (2019) External validation of prognostic models for preeclampsia in a Dutch multicenter prospective cohort, Hypertension in Pregnancy, 38:2, 78-88, DOI: 10.1080/10641955.2019.1584210

To link to this article: https://doi.org/10.1080/10641955.2019.1584210

HYPERTENSION IN PREGNANCY
2019, VOL. 38, NO. 2, 78–88
https://doi.org/10.1080/10641955.2019.1584210

Taylor & Francis
Taylor & Francis Group

External validation of prognostic models for preeclampsia in a Dutch multicenter prospective cohort

Marije Lamain-de Ruiter [a], Anneke Kwee[a], Christiana A. Naaktgeboren[b], Rebecca D. Louhanepessy[c], Inge De Groot[d], Inge M. Evers[e], Floris Groenendaal[f], Yolanda R. Hering[g], Anjoke J. M. Huisjes[h], Cornel Kirpestein[i], Wilma M. Monincx[j], Peter C. J. I. Schielen[k], Annewil Van 'T Zelfde[l], Charlotte M. Van Oirschot[m], Simone A. Vankan-Buitelaar[n], Mariska A. A. W. Vonk[o], Therese A. Wiegers[p], Joost J. Zwart[q], Karel G. M. Moons[b], Arie Franx [a], and Maria P. H. Koster[a,r]

[a]Department of Obstetrics, Division Woman and Baby, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [b]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [c]Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands; [d]Livive, Center for Obstetrics, Tilburg, The Netherlands; [e]Department of Obstetrics, Meander Medical Center, Amersfoort, The Netherlands; [f]Department of Neonatology, Division Woman and Baby, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [g]Department of Obstetrics, Zuwe Hofpoort Hospital, Woerden, The Netherlands; [h]Department of Obstetrics, Gelre Hospital, Apeldoorn, The Netherlands; [i]Department of Obstetrics, Hospital Rivierenland, Tiel, The Netherlands; [j]Department of Obstetrics, St. Antonius Hospital, Nieuwegein, The Netherland; [k]Center for Infectious Diseases Research, Diagnostics and Screening (IDS), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; [l]Midwifery practice Verloskundigen Amersfoort, Amersfoort, The Netherlands; [m]Department of Obstetrics, St Elisabeth Hospital, Tilburg, The Netherlands; [n]Midwifery practice GCM, Maarssen, The Netherlands; [o]Midwifery practice Het Wonder, Houten, The Netherlands; [p]Netherlands Institute for health services research (NIVEL), Utrecht, The Netherlands; [q]Department of Obstetrics, Deventer Hospital, Deventer, The Netherlands; [r]Department of Obstetrics and Gynecology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, the Netherlands

ABSTRACT

Objective: To perform an external validation of all published prognostic models for first-trimester prediction of the risk of developing preeclampsia (PE).

Methods: Women <14 weeks of pregnancy were recruited in the Netherlands. All systematically identified prognostic models for PE that contained predictors commonly available were eligible for external validation.

Results: 3,736 women were included; 87 (2.3%) developed PE. Calibration was poor due to overestimation. Discrimination of 9 models for LO-PE ranged from 0.58 to 0.71 and of 9 models for all PE from 0.55 to 0.75.

Conclusion: Only a few easily applicable prognostic models for all PE showed discrimination above 0.70, which is considered an acceptable performance.

ARTICLE HISTORY
Received 29 June 2018
Accepted 13 February 2019

KEYWORDS
First trimester; preeclampsia; external validation; prognostic model

Introduction

Preeclampsia (PE) is one of the leading causes of maternal and perinatal mortality and morbidity (1). PE complicates approximately 2–5% of all pregnancies (2) and is characterized by new onset of hypertension and proteinuria after 20 weeks of pregnancy (3). Preventive measures, like prescription of calcium and low-dose aspirin, started during the first trimester, have been proven to prevent PE (4,5). Currently, the administration of those preventive measures is based on the presence of risk factors known for PE, such as history of PE or chronic hypertension (6,7). However, combining risk factors in prognostic models often allows for better risk assessment compared to single risk predictors.

To date, numerous multivariable prognostic models have been developed to predict PE (8–10). Recent quality assessments of first trimester prognostic models for PE have shown that methodological flaws are frequently present (9,10). These flaws, such as low number of events and inferior selection methods of risk predictors, may limit the validity and reproducibility of prognostic models. Moreover, when used in routine antenatal care, their performance may be worse compared to the development setting. This emphasizes the importance of external validation of prognostic models in independent datasets to assess their clinical value. Up until now, only a few prognostic models for PE have been externally validated (10). In order to acquire a fair comparison of their predictive

CONTACT Maria P. H. Koster m.p.h.koster@erasmusmc.nl Department of Obstetrics & Gynaecology, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000CA Rotterdam, the Netherlands
Supplementary data for this article are accessed here.

accuracy, the aim of our study was to perform an external validation by examining the performance of published first trimester prognostic models for PE in one independent cohort. Models consisting of commonly available predictors were selected in order to validate models that are easily applicable in clinical practice with only limited costs, even in low resource countries.

Methods

Study population

From December 2012 through January 2014 pregnant women were recruited in the RESPECT cohort (Risk Estimation for PrEgnancy Complications to provide Tailored care). A detailed description of the design and participants of this study has been described previously (11). In short, all consecutive women were included at their initial prenatal visit (<14 weeks of pregnancy) in 31 independent midwifery practices (primary care) and six hospitals (secondary/tertiary care) in the central region of the Netherlands. During the course of their pregnancy participants received routine antenatal care according to Dutch clinical guidelines. In the Netherlands, pregnant women were considered at high risk of developing PE when they had a history of PE, a history of intrauterine growth restriction (requiring childbirth prior to 34 weeks of pregnancy), or a history of a chronic condition leading to placental insufficiency (e.g. severe renal dysfunction or systematic lupus erythematosus). Only these women were eligible for administration of aspirin, resulting in 1 woman using aspirin during pregnancy. Data on women who miscarried before 16 weeks were excluded from the analysis.

This study was approved by the medical ethics committee of the University Medical Center Utrecht (protocol number 12–432/C) and written informed consent was obtained from all participants. Results have been reported conform to the TRIPOD statement (12).

Predictor assessment

At the initial study visit in the first trimester of pregnancy, several predictors were measured, such as maternal age, body mass index, and blood pressure. Between 9 and 14 weeks of gestation blood was withdrawn to measure the biochemical serum markers pregnancy-associated plasma protein-A (PAPP-A) and placental growth factor (PlGF). Maternal characteristics and medical and obstetrical history were obtained through a self-administered questionnaire in the first trimester of pregnancy. A full description of all predictor and marker definitions can be found in Appendix A. Distribution of predictors among original studies was not reported, because this information was often lacking.

Outcome assessment

PE was defined according to definition of the International Society for the Study of Hypertension in Pregnancy (3). PE was diagnosed if the diastolic blood pressure was 90 mmHg or higher on at least two separate occasions after 20 weeks of gestation in previously normotensive women combined with the presence of proteinuria of 300 mg or more during 24 h. All cases of PE as "all PE". PE cases requiring childbirth after 34 weeks of gestation were defined as "late-onset PE" (LO-PE). Cases requiring childbirth before 34 weeks of gestation were defined as "early onset PE" (EO-PE).

Selection of prognostic models for external validation

For the selection of prognostic models, we have updated a systematic review on models for several obstetric complications previously published by Kleinrouweler *et al.* (8). Medline and Embase were searched from 1 January 2012 till 23 December 2014. A combination of terms for first trimester of pregnancy, PE and a validated search strategy for prediction modelling studies was used. The exact search details and a short summary of this systematic review are provided in Appendix B. We chose to limit our search to models that consist of commonly available predictors, which are therefore widely applicable in clinical practice with only limited costs, even in low resource countries.

Prognostic models predicting PE (all PE or LO-PE) based on easily measurable predictors, available before 14 weeks of pregnancy, were eligible. Models including the commonly used biomarkers PAPP-A and PlGF were also eligible.

For the purpose of external validation, the exact definition of the predictors included in the model, how the predictors were measured, and the exact prediction equation were retrieved from the original publications. If information on predictor definition, intercepts or coefficients were missing, authors were contacted by email ($n = 13$). Two authors responded and provided this information. Due to the low incidence of EO-PE in the RESPECT cohort we had to exclude prognostic models for EO-PE. Eventually, 18 prognostic models for all PE or LO-PE remained for external validation in the current study (13–29). A description of the exact predictors and equations used for external validation is

reported in Appendix A and Appendix C, respectively. Distributions of predictors in the original studies were not reported as this information was often lacking in the original publications.

Statistical analysis

Missing values on predictors or outcome in the validation cohort were imputed by multiple imputation, based on the assumption that this data was not missing completely at random, as can directly be concluded from Table 1 (30). Continuous variables were compared using t-tests or Mann–Whitney U tests depending on their distributions, while categorical variables were compared using X^2 test. All possible predictors and outcomes were used in the imputation model and ten imputations were performed. Results shown are the results after multiple imputations, unless otherwise specified. All analyses were carried out on each of the multiple imputed datasets and Rubin's rules were used to combine the results into summary estimates (31). Analyses were performed using the mice and rms packages of R-3.1 for windows (http://cran.r-project.org).

First, the predicted probabilities for each participant in the RESPECT cohort were calculated-based on the exact prognostic models as published, the "original" results. This can only be performed when the full prediction rule, including its intercept, is available (Appendix C).

Second, "logistic calibration" was performed to allow for a fair comparison of the models. For this adjustment, the linear predictor is used as the only covariate and a updated calibration slope and intercept were calculated (32,33). Results are shown as "recalibrated".

Third, to assess whether the results were not merely the result of a poor fit on our population, the prognostic models were completely "refitted" to our population. This way we were able to quantify each model's maximal predictive accuracy which we could compare to the results after validation of the originally published models (34). This results in a new intercept and new regression coefficients for each prognostic model. Results are shown as "refitted".

The performance of each prognostic model for PE was assessed in terms of calibration and discrimination. Calibration of "original" and "recalibrated" models was observed using calibration plots. A calibration plot compares the predicted probabilities of PE for each individual with the observed outcome. The predicted probabilities equal the observed proportions for all groups, normally 10, when a model is well calibrated. A calibration plot has an intercept of 0 and a slope of 1 and all groups ideally fit close to this diagonal line. The updated calibration intercept and slope of the linear prediction after recalibration were used to assess model estimation and overfitting. Overestimation is probably present when the calibration intercept is less than 0, whereas underestimation is probably present when the calibration intercept is greater than 0. Overfitting of the original prognostic model is indicated by a calibration slope of less 1 (35).

Discrimination was assessed using the area under the receiver operating characteristic curve (AUC) (36). The AUC is used to verify whether participants with a higher predicted risk for PE are indeed more likely to develop PE. An AUC of 0.50 offers no statistical improvement over a random guess, whereas an AUC of 1.00 would mean perfect prediction for all participants.

Since a history of PE is an important predictor in most prognostic models, a subgroup analysis was performed in nulliparous women. Discrimination and calibration were re-assessed in this subgroup.

Finally, a table was constructed with the distribution of women with and without PE among several predicted risk categories, based on the prognostic models that showed good calibration (slope > 0.80). The current NICE (National Institute for Health and Care Excellence) guideline for risk reduction of PE was applied to our cohort for comparison of performance with the prognostic models (7). This guideline advises to prescribe aspirin in case women have one or more high-risk factors or two or more moderate-risk factors.

Results

RESPECT cohort

Our validation cohort included 3,736 pregnant women of whom 1,662 (44%) were nulliparous. Other baseline characteristics of our study population are shown in Table 1. A total of 87 (2.3%) women developed PE of whom 71 (1.9%) had LO-PE and 16 (0.4%) EO-PE. Superimprosed preeclampsia occurred in 2 women, both had LO-PE. In the nulliparous subgroup 65 (4.0%) women developed PE of whom 51 (3.1%) had LO-PE and 14 (0.9%) EO-PE.

Calibration and discrimination

Table 2 summarizes all predictors that were included in the prognostic models and measured in our validation cohort. Original models for all and LO-PE were applied to the validation cohort when the original publications provided the full prediction rule, which was the case for 7 out of 9 all PE models and for 6 out of 9 LO-PE models. Calibration of prognostic models for all PE and LO-PE was poor. Most original models for all PE, as well as LO-PE, seemed to overestimate the risk of PE, as can be seen

Table 1. Baseline characteristics stratified per variable that were present for imputation [11].

Characteristic	Cases with missing value n(%)	Complete questionnaires (n = 2614)	Cases with ≥ 1 missing questionnaire (n = 1122)	*p* value	Overall RESPECT cohort (n = 3736)
Age, yrs	168 (4.5%)	30.9 ± 4.3	30.7 ± 3.9	0.32	30.8 ± 4.2
BMI pre-pregnancy, kg/m^2	46 (1.2%)	23.3 [21.2, 26.3]	23.0 [20.9, 25.9]	0.02*	23.2 [21.1, 26.2]
BMI, kg/m^2	184 (4.9%)	23.8 [21.6, 26.9]	23.4 [21.3, 26.3]	0.01*	23.7 [21.5, 26.7]
Systolic BP, mmHg	65 (1.7%)	114 ± 12	115 ± 12	0.55	114 ± 12
Diastolic BP, mmHg	64 (1.7%)	67 ± 8	67 ± 8	0.10	67 ± 8
Etnicity,	736 (19.7%)	1671 (89.0%)	1068 (95.2%)	<0.001*	3398 (91.0%)
– White		18 (1.0%)	2 (0.2%)		31 (0.8%)
– African		30 (1.6%)	11 (1.0%)		54 (1.4%)
– Asian		44 (2.3%)	15 (1.3%)		77 (2.1%)
– Mixed		115 (4.4%)	26 (2.3%)		176 (4.7%)
– Other					
Education,	225 (6.0%)	199 (7.6%)	52 (4.6%)	0.001*	272 (7.3%)
– Low		826 (31.6%)	364 (32.4%)		1277 (34.2%)
– Middle		1364 (57.1%)	706 (62.9%)		2187 (58.6%)
– High					
Smoking during pregnancy	0 (0%)	260 (9.9%)	73 (6.5%)	0.001*	336 (9.0%)
History of chronic hypertension	1 (0.0%)	45 (1.7%)	15 (1.3%)	0.47	60 (1.6%)
Mother with PE	1 (0.0%)	75 (2.9%)	41 (3.7%)	0.24	116 (3.1%)
Method of conception	30 (0.8%)	2407 (93.2%)	1035 (92.2%)	0.20	3469 (92.8%)
– Spontaneous		61 (2.4%)	38 (3.4%)		100 (2.7%)
– Ovulation drugs		82 (3.2%)	28 (2.5%)		111 (3.0%)
– IVF					
Nulliparous	4 (0.0%)	1149 (44.0%)	510 (45.5%)	0.44	1662 (44.5%)
History of PE	0 (0.0%)	86 (3.3%)	35 (3.1%)	0.87	121 (3.2%)
History of IUGR (<10th perc)	0 (0.0%)	93 (3.6%)	39 (3.5%)	0.98	132 (3.5%)
Recurrent miscarriages (≥2)	4 (0.0%)	174 (6.7%)	59 (5.3%)	0.12	233 (6.2%)
History of fetal death	0 (0.0%)	58 (2.2%)	16 (1.4%)	0.14	74 (2.0%)
PE	265 (7.1%)	56 (2.4%)	22 (2.0%)	0.51	87 (2.3%)
– Late onset		45 (1.9%)	20 (1.8%)	0.89	71 (1.9%)
Gestational age at birth, days	344 (9.2%)	280 [273, 285]	280 [274, 286]	0.33	280 [273, 285]
Sex, male	360 (9.6%)	1159 (50.7%)	570 (52.3%)	0.41	1909 (51.1%)
Birth weight, g	374 (10.0%)	3503 [3200, 3855]	3540 [3219, 3880]	0.11	3520 [3190, 3875]
– percentile		54 (30, 77)	57 (32, 80)	0.07	55 (30, 79)
– <10th percentile		138 (6.4%)	63 (5.9%)	0.62	270 (7.2%)

yrs, years; BMI, body mass index; BP, blood pressure; PE, preeclampsia; IVF, in vitro fertilization; IUGR, intra uterine growth restriction; Data are n, n(%), mean±SD, or median [IQR]. The column "overall RESPECT cohort" includes imputed data for those with missing values.
*Significant at the P < 0.05 level.

Table 2. Summary of predictors per model.

	Predictors	Maternal age	Weight	Length	BMI, prepregnancy	BMI	Blood pressure	History of PE	History of IUGR	Parity	Family history of PE	History of chronic hypertension	History of DM t 1/2	History of thrombo-embolic proces	Ethnicity	Method of conception	Smoking	Education level	a priori	PAPP-A MoM	PIGF MoM	Total
All preeclampsia	Baschat 2014						x	x		x		x								x		5
	Giguere 2014				x		x															2
	Goetzinger 2010					x						x	x		x					x		5
	Goetzinger 2013					x		x				x	x							x		5
	Myatt 2012					x	x								x			x				4
	Odibo 2011					x						x			x					x		4
	Plasencia 2007					x		x		x	x				x							5
	Poon 2008					x	x	x		x	x				x				x			7
	Syngelaki 2011	x				x		x		x	x	x	x		x	x	x					10
	Total all preeclampsia	**1**	**0**	**0**	**1**	**7**	**4**	**5**	**0**	**4**	**3**	**5**	**3**	**0**	**6**	**1**	**1**	**1**	**1**	**4**	**0**	**5**
Late onset preeclampsia	Akolekar 2011		x	x				x		x		x			x							6
	Crovetto 2014a					x		x				x			x		x					5
	Crovetto 2014b					x		x	x	x												4
	Kuc 2013	x	x							x												3
	Kuc 2014	x				x				x												
	Plasencia 2007					x		x		x	x				x							5
	Poon 2009					x		x		x	x				x					x		6
	Poon 2010	x				x		x		x	x				x							6
	Scazzochio 2013					x				x		x	x	x								5
	Total late onset preeclampsia	**3**	**2**	**1**	**0**	**7**	**0**	**6**	**1**	**8**	**3**	**3**	**1**	**1**	**5**	**0**	**1**	**0**	**0**	**1**	**0**	**5**

All predictors are measured in the first trimester at the initial prenatal visit, unless otherwise specified. Modifications to the original predictors are described in Appendix A. Abbreviations: BMI, body mass index; PE, preeclampsia; IUGR, intra-uterine growth restriction; DM, diabetes mellitus; PAPP-A, pregnancy-associated plasma protein-A; PlGF, placental growth factor.

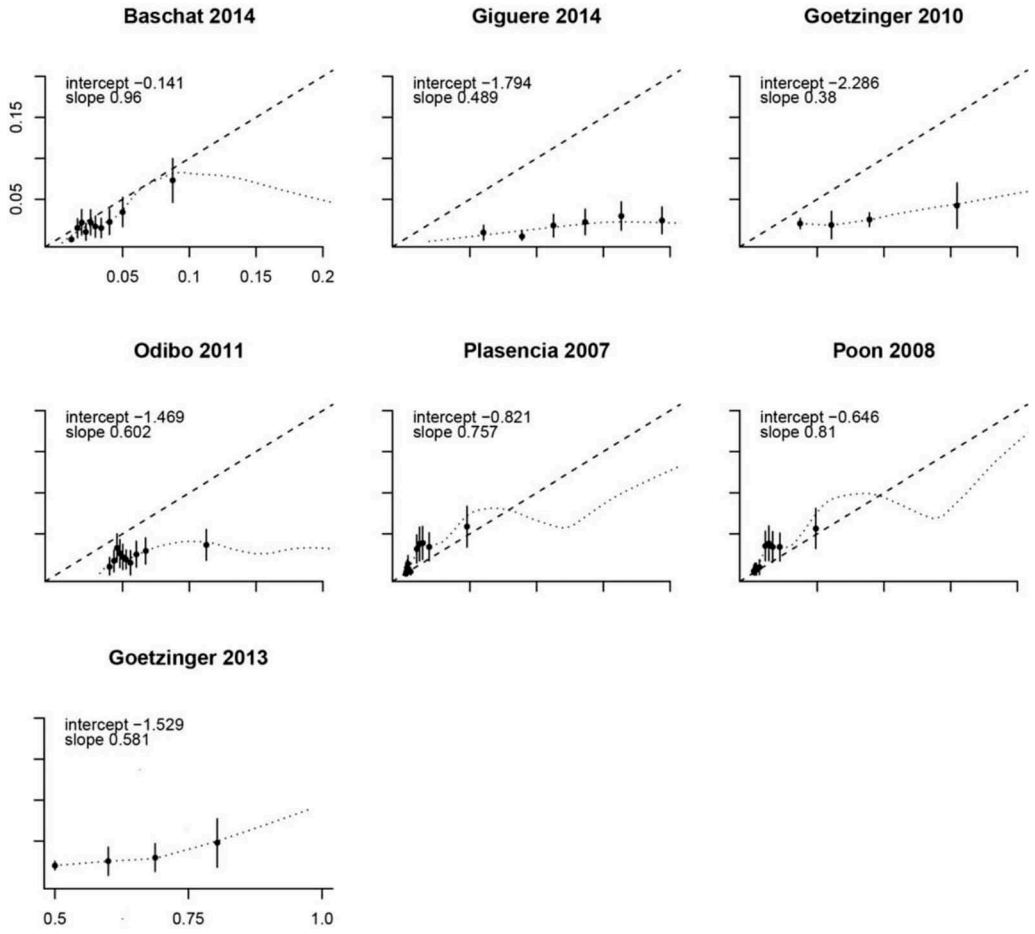

Figure 1. Calibration plots of original prognostic models.

In case of perfect calibration all groups of predicted probabilities fit close to the diagonal line, corresponding with an intercept of 0 and a slope of 1 for the calibration plot. Vertical lines in grouped observed represent 95% confidence intervals.

by the predicted risks which were higher than the observed risks (Figure 1). Recalibration yielded some improvement, but except for one model risk overestimation (intercept less than 0) was still present (Figure 2). The models by Baschat *et al.* 2014, Syngelaki *et al.* 2011 and Poon *et al.* 2008 showed the least overfitting (calibration slopes above 0.80) for all PE. For LO-PE, Akokelar *et al.* 2011 and Kuc *et al.* 2014 both had a calibration slope above 0.80.

In terms of discrimination, the AUC ranged from 0.55 to 0.72 for all PE (after recalibration) (Table 3). The models by Plasencia *et al.* 2007, Poon *et al.* 2008 and Syngelaki *et al.* 2011 had the highest AUC, all above 0.70. When the models were completely refitted to the study population, the AUC showed only marginal improvement indicating that the discrimination of the original models could not be improved by complete refitting on our data set (Appendix D, figure A).

Models that included the history of PE had a higher AUC than those that did not (Appendix D, figure C). For LO-PE models the AUC tended to be lower and ranged from 0.58 to 0.71 (Appendix D, figure A). The models by Akolekar *et al.* 2011, Poon *et al.* 2009 and Poon *et al.* 2010 all had an AUC of 0.70 or higher. Refitted models again showed a slight improvement of the AUC. Applying the prognostic models for all and late-onset PE to a subgroup of nulliparous women yielded poor discrimination (0.50 to 0.63 and 0.48 to 0.59, respectively) (Appendix D, figure B).

Predicted risk categories

Table 4 shows the number of women who did and did not develop PE, stratified by predicted risk category for the refitted prognostic models that were properly calibrated for all PE (three models) and LO-PE (two models). For all PE models, most women who developed PE were in the highest categories and only few of these women were in the lowest risk category. For example, a predicted risk threshold of ≥5% would correctly classify 40% of women who developed PE as high-risk (sensitivity) and 90% of women who did not develop PE as low-

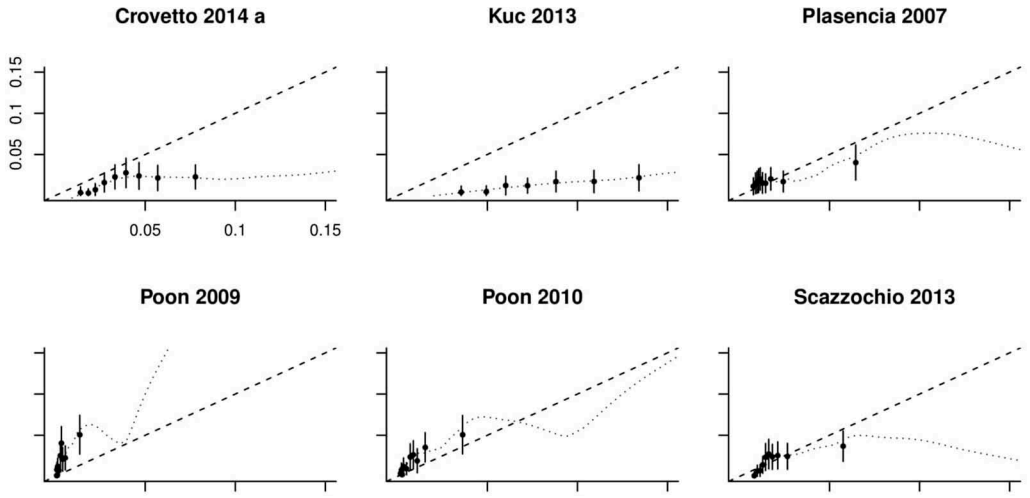

Figure 2. Calibration plots of recalibrated prognostic models.

In case of perfect calibration all groups of predicted probabilities fit close to the diagonal line, corresponding with an intercept of 0 and a slope of 1 for the calibration plot. Vertical lines in grouped observed represent 95% confidence intervals.

risk (specificity) for the model of Baschat *et al.* 2014. For LO-PE similar results were observed.

In comparison, applying the current NICE guideline for risk reduction of PE in our cohort would classify 10% of all women at risk for PE with a sensitivity of 28% and specificity of 90% (7).

Discussion

Main findings

In this study, a comparison of nine first trimester prognostic models for all PE and nine prognostic models for LO-PE showed discrimination between 0.55 and 0.75 and moderate calibration with slopes of 0.13 to 1.19. Three models for all PE (Plasencia *et al.* 2007, Poon *et al.* 2008 and Syngelaki *et al.* 2011) and three models for LO-PE (Akolekar *et al.* 2011, Poon *et al.* 2009 and Poon *et al.* 2010) had an AUC above 0.70 which was not much improved after completely refitting the models. The most common predictors in these models were body mass index (BMI), parity, history of PE, history of chronic hypertension and ethnicity. Performance in a subgroup of nulliparous women yielded discrimination below 0.65 for all models, probably because the history of PE is a strong predictor in most models. Overall, with a predicted risk cut-off of ≥5% approximately 40% of all women who will develop PE can be identified.

Strengths and limitations

Our external validation study is one of the first studies that compares a large number of published first trimester prognostic models for PE in one single independent cohort. This prospective multicenter cohort consisted of both high- and low-risk women, strengthening the generalizability of our results.

However, some limitations need to be considered. We restricted this external validation to published prognostic models that included maternal characteristics and/or two commonly used serum biomarkers (PAPP-A and PlGF). Thereby, promising prognostic models for PE including other biochemical markers and/or uterine artery Doppler assessment might have been missed. As it appeared especially hard to predict PE in nulliparous women, a specialized prognostic approach (combination of maternal characteristics, placental markers, and vascular markers) would be more in line with the multifaceted origin of this major pregnancy-syndrome and might improve discrimination of prognostic models for PE in nulliparous women. Also, we focused on all and LO-PE, but an external validation of published models for the prediction of EO-PE is recommended. However, due to its low incidence, this can probably best be performed by combining datasets in an individual patient data meta-analysis. Another advantage of such a study is that it would probably result in a more ethnically diverse cohort.

Interpretation

Risk factors that are most strongly associated with PE were often used in prognostic models for PE, e.g. history of PE and BMI (37). Al-Rubaie *et al.* provided an overview of the performance of "simple" risk models for PE as reported in the original publications and showed a wide variety of the discriminative ability of these models (10). More recently, the ASPRE trial

Table 3. Discriminative ability of all prognostic models in the external validation.

Prognostic model	AUC Development	AUC Recalibrated	AUC Refitted	AUC Nulliparous Recalibrated	AUC Nulliparous Refitted
ALL PE					
Baschat '14	0.82 [0.78 to 0.86]	0.68 [0.61 to 0.74]	0.76 [0.71 to 0.81]	0.63 [0.55 to 0.71]	0.64 [0.56 to 0.72]
Giguere '14	0.75 [0.69 to 0.81]	0.63 [0.57 to 0.69]	0.64 [0.58 to 0.71]	0.61 [0.54 to 0.69]	0.63 [0.55 to 0.71]
Goetzinger '10	0.70 [0.65 to 0.72]	0.55 [0.48 to 0.61]	0.55 [0.49 to 0.61]	0.50 [0.43 to 0.57]	0.50 [0.43 to 0.57]
Goetzinger '13	0.76 [0.69 to 0.83]	0.56 [0.50 to 0.61]	0.56 [0.50 to 0.61]	0.52 [0.46 to 0.57]	0.52 [0.46 to 0.57]
Myatt '12	0.65 [0.61 to 0.69]	0.64 [0.58 to 0.70]	0.64 [0.58 to 0.70]	0.61 [0.53 to 0.68]	0.62 [0.54 to 0.69]
Odibo '11	0.77 [0.63 to 0.81]	0.56 [0.49 to 0.62]	0.57 [0.50 to 0.63]	0.52 [0.45 to 0.59]	0.53 [0.46 to 0.61]
Plasencia '07	0.81 [0.80 to 0.82]	0.72 [0.67 to 0.77]	0.73 [0.68 to 0.78]	0.53 [0.45 to 0.59]	0.54 [0.47 to 0.62]
Poon '08	0.85 [NR]	0.71 [0.66 to 0.76]	0.76 [0.71 to 0.81]	0.51 [0.43 to 0.59]	0.63 [0.55 to 0.71]
Syngelaki '11	NR	0.72 [0.67 to 0.78]	0.75 [0.70 to 0.80]	0.55 [0.47 to 0.63]	0.59 [0.51 to 0.67]
LATE ONSET PE					
Akolekar '11	NR	0.71 [0.65 to 0.77]	0.72 [0.66 to .78]	0.53 [0.45 to 0.62]	0.54 [0.45 to 0.62]
Crovetto '14a	0.72 [0.69 to 0.76]	0.66 [0.59 to 0.72]	0.73 [0.67 to 0.79]	0.48 [0.40 to 0.57]	0.57 [0.48 to 0.66]
Crovetto '14b	0.75 [0.67 to 0.82]	0.58 [0.51 to 0.65]	0.58 [0.50 to 0.65]	0.53 [0.45 to 0.62]	0.53 [0.45 to 0.62]
Kuc '13	NR	0.66 [0.60 to 0.73]	0.68 [0.62 to 0.74]	0.53 [0.45 to 0.61]	0.53 [0.45 to 0.61]
Kuc '14	0.79 NR	0.67 [0.61 to 0.74]	0.68 [0.62 to 0.74]	0.59 [0.51 to 0.68]	0.60 [0.52 to 0.68]
Plasencia '07	0.80 [0.79 to 0.81]	0.58 [0.51 to 0.65]	0.60 [0.53 to 0.67]	0.53 [0.45 to 0.61]	0.55 [0.47 to 0.63]
Poon '09	0.79 [0.78 to 0.80]	0.70 [0.64 to 0.76]	0.73 [0.67 to 0.79]	0.52 [0.43 to 0.60]	0.57 [0.49 to 0.66]
Poon '10	0.80 [0.76 to 0.83]	0.70 [0.64 to 0.77]	0.73 [0.67 to 0.79]	0.53 [0.45 to 0.61]	0.55 [0.47 to 0.63]
Scazzochio '13	0.71 [0.66 to 0.76]	0.67 [0.61 to 0.73]	0.69 [0.62 to 0.75]	0.54 [0.45 to 0.62]	0.57 [0.48 to 0.66]

The AUC "development" shows the AUC as reported in the original publication if available. The AUC "recalibrated" shows the AUC per model, recalibrated to the RESPECT cohort. The AUC "refitted" shows the AUC per model after complete refitting of the prognostic model to the RESPECT cohort. The AUC "Nulliparous" shows the AUC per model when applied to a subgroup of only nulliparous.
Abbreviations: AUC, area under the receiver operating characteristic curve; PE, preeclampsia; NR, not reported.
Data are presented in mean [95% confidence interval].

showed a detection rate of 38.3% for term PE (38). Our external validation study confirmed these findings, especially the poor performance of prognostic models for LO-PE (detection rate 36%, Table 4).

Some prognostic models that were validated in our study have been externally validated before. For example, the LO-PE model by Scazzochio *et al.* with an original AUC of 0.71 showed an AUC of 0.69 in a previous validation study and was 0.67 in our validation study (39). For the model for LO-PE by Plasencia *et al.* larger differences were observed. Their original development study showed an AUC of 0.80, where validation studies showed a significantly lower AUC: 0.72, 0.65 and 0.58 in our validation study (40,41). Since the calibration slope for the model by Plascensia *et al.* was low, the large variation in AUCs may very well be the effect of model overfitting.

The main benefit of first-trimester prognostic models for PE is that they help to guide individualized planning of antenatal care, and to decide on the prescription of preventive measures such as low-dose aspirin and calcium supplementation. However, before implementing prognostic models, it is important to assess their true value in an external validation study. As only a few models showed proper calibration and discrimination was limited to 0.76 at most, the applicability for clinical practice, especially for the nulliparous subgroup, may be limited. On the other hand, the discriminative ability of the prognostic models with the highest AUC outperform current single risk factor strategies. For example, the model by Baschat *et al.* 2014 can

Table 4. Pregnancy outcome per predicted risk category.

All PE	Baschat 2014						Poon 2008						Syngelaki 2011					
Predicted risk %	no PE	%	FPR	PE	%	DR	no PE	%	FPR	PE	%	DR	no PE	%	FPR	PE	%	DR
>5.0	369	10	10	35	40	40	309	8	8	21	24	24	398	11	11	33	38	38
4.5 to < 5.0	135	4	14	3	4	44	115	3	12	6	6	31	119	3	14	6	7	45
4.0 to < 4.5	195	5	19	10	11	55	214	6	17	7	8	39	173	5	19	5	6	50
3.5 to < 4.0	244	7	26	5	6	61	397	11	28	17	19	58	211	6	25	6	7	57
3.0 to < 3.5	264	7	33	7	8	69	509	14	42	20	23	82	256	7	32	10	12	69
2.5 to < 3.0	230	6	39	8	10	79	128	4	46	3	4	85	251	7	39	8	9	79
2.0 to < 2.5	188	5	45	6	6	85	5	0	46	0	0	85	201	5	44	5	5	84
1.5 to < 2.0	96	3	47	0	0	85	6	0	46	1	1	86	104	3	47	2	3	86
1.0 to < 1.5	170	5	52	2	2	88	75	2	48	0	0	86	186	5	52	1	2	88
0.5 to < 1.0	1105	30	82	9	10	98	1669	46	94	12	13	100	1068	29	81	9	10	98
<0.5	653	18	100	2	2	100	222	6	100	0	0	100	682	19	100	2	2	100
Late onset PE	**Akolekar 2011**						**Kuc 2014**											
Predicted risk %	no PE	%	FPR	PE	%	DR	no PE	%	FPR	PE	%	DR						
>5.0	349	10	10	25	36	36	108	3	3	5	7	7						
4.5 to < 5.0	92	3	12	4	6	41	32	1	4	5	7	14						
4.0 to < 4.5	146	4	16	3	5	46	58	2	5	1	2	15						
3.5 to < 4.0	217	6	22	6	8	54	66	2	7	4	5	20						
3.0 to < 3.5	442	12	34	9	13	67	132	4	11	5	6	27						
2.5 to < 3.0	407	11	45	11	15	82	233	6	17	7	10	36						
2.0 to < 2.5	37	1	46	1	1	83	400	11	28	14	20	57						
1.5 to < 2.0	35	1	47	0	0	83	793	22	50	13	18	74						
1.0 to < 1.5	91	2	49	1	2	85	1614	44	94	18	26	100						
0.5 to < 1.0	1481	40	90	10	14	99	231	6	100	0	0	100						
<0.5	371	10	100	1	1	100	0	0	100	0	0	100						

PE, preeclampsia; FPR, false positive rate; DR, detection rate

correctly identify 40% of all women who will develop PE at a predicted risk cut-off point of 5% [Table 4]. Moreover, the findings of the ASPRE trial shows that the incidence of EO-PE is more than halved when low-dose aspirin is prescribed to women who are detected to be at high risk of developing PE by a prognostic model (42). Therefore, an cost-effectiveness analysis on the use of prognostic models for PE to guide the decision-making of preventive measures would be the next step to provide more insight into the harms and benefits compared to current single risk factor strategies.

Although there might be room for improvement of current prognostic models for all PE, when clinicians want to make use of a model, we recommend choosing one of the models for all PE with the highest AUC. Based on their performance in our external validation, development or use of first-trimester models predicting LO-PE using only commonly available predictors is not recommended, especially not for nulliparous women.

Acknowledgments

We would like to thank all pregnant women who participated in the RESPECT study. A special thanks to the midwifery practices and hospitals in our regional consortium (GCMN) for their help in recruiting participants and technicians (I. Belmouden, G. Diependaal, M. Jonker, P. Turion, S. Verhoef) of the National Institute for Public Health and the Environment (RIVM) and U. Koster for their help in analyzing the biochemical serum markers.

Authors' contributions

MPHK, AK, AF, KGMM and the RESPECT study group (IdG, IME, FG, YRH, AJMH, CK, WMM, PCJIS, Av'tZ, CMvO, SAVB, MAAWV, TAW, JZ) had the original idea for the study and were involved in writing the original study protocol. The RESPECT study group and MLdR were involved in data collection. MLdR and RDL carried out the systematic review. CAN and MLdR performed data analysis. MLdR, CAN, and MPHK wrote the first draft of the manuscript, which was subsequently revised by AF, AK, and KGMM. All authors participated in the final approval of the manuscript. MPHK and AF are the guarantors of this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by The Netherlands Organization for Health Research and Development under Grant [project nr 209020004].

Ethical approval

This study was approved by the medical ethics committee of the University Medical Center Utrecht (protocol number 12-432/C) and written informed consent was obtained from all participants.

ORCID

Marije Lamain-de Ruiter http://orcid.org/0000-0002-5616-5686
Arie Franx http://orcid.org/0000-0001-8801-5546

References

[1] Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol. 2009;33:130–137.
[2] Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. BMJ. 2013;347:f6564.
[3] Brown MA, Lindheimer MD, de Swiet M, et al. The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the international society for the study of hypertension in pregnancy (ISSHP). Hypertens Pregnancy. 2001;20:IX–XIV.
[4] Roberge S, Nicolaides KH, Demers S, et al. Prevention of perinatal death and adverse perinatal outcome using low-dose aspirin: a meta-analysis. Ultrasound Obs Gynecol. 2013;41:491–499.
[5] Hofmeyr JG, Lawrie TA, Atallah AN, et al. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database Syst Rev. 2018;6:CD001059.
[6] 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.
[7] National Institute for Health and Care Excellence. Hypertension in pregnancy. NICE guideline CG107; 2011.
[8] Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol. 2016;214:79–90.e36.
[9] Brunelli VB, Prefumo F. Quality of first trimester risk prediction models for pre-eclampsia: a systematic review. BJOG. 2015;122:904–914.
[10] Al-Rubaie Z, Askie L, Ray J, et al. The performance of risk prediction models for pre-eclampsia using routinely collected maternal characteristics and comparison with models that include specialised tests and with clinical guideline decision rules: a systematic review. BJOG. 2016;123(9):1441–1452.
[11] Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. BMJ. 2016;354:i4338.
[12] Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1.
[13] Akolekar R, Syngelaki A, Sarquis R, et al. Prediction of early, intermediate and late pre-eclampsia from maternal factors, biophysical and biochemical markers at 11–13 weeks. Prenat Diagn. 2011;3:66–74.
[14] Baschat AA, Magder LS, Doyle LE, et al. Prediction of preeclampsia utilizing the first trimester screening examination. Am J Obstet Gynecol. 2014;211:514.e1–7.
[15] Crovetto F, Figueras F, Triunfo S, et al. First trimester screening for early and late preeclampsia based on maternal characteristics, biophysical parameters, and angiogenic factors. Prenat Diagn. 2015;35:183–191.
[16] Crovetto F, Figueras F, Triunfo S, et al. Added value of angiogenic factors for the prediction of early and late preeclampsia in the first trimester of pregnancy. Fetal Diagn Ther. 2014;35:258–266.
[17] Giguere Y, Masse J, Theriault S, et al. Screening for pre-eclampsia early in pregnancy: performance of a multivariable model combining clinical characteristics and biochemical markers. BJOG. 2015;122:402–410.
[18] Kuc S, Koster MP, Pennings JL, et al. Metabolomics profiling for identification of novel potential markers in early prediction of preeclampsia. PLoSOne. 2014;9:e98540.
[19] Kuc S, Koster MPH, Franx A, et al. Maternal characteristics, mean arterial pressure and serum markers in early prediction of preeclampsia. PLoS One. 2013;8:e63546.
[20] Myatt L, Clifton RG, Roberts JM, et al. First-trimester prediction of preeclampsia in nulliparous women at low risk. Obstet Gynecol. 2012;119:1234–1242.
[21] Odibo AO, Zhong Y, Goetzinger KR, et al. First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. Placenta. 2011;32:598–602.
[22] Plasencia W, Maiz N, Bonino S, et al. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. Ultrasound Obs. 2007;30:742–749.
[23] Poon LC, Kametas NA, Chelemen T, et al. Maternal risk factors for hypertensive disorders in pregnancy: a multivariate approach. JHumHypertens. 2010;24:104–110.
[24] Lc P, Maiz N, Valencia C, et al. First-trimester maternal serum pregnancy-associated plasma protein-A and pre-eclampsia. Ultrasound Obs. 2009;33:23–33.
[25] Poon LCY, Kametas NA, Pandeva I, et al. Mean arterial pressure at 11(+0) to 13(+6) weeks in the prediction of preeclampsia. Hypertension. 2008;51:1027–1033.
[26] Scazzocchio E, Figueras F, Crispi F, et al. Performance of a first-trimester screening of preeclampsia in a routine care low-risk setting. Am J Obstet Gynecol. 2013;208:203.e1–203.e10.
[27] Goetzinger KR, Singla A, Gerkowicz S, et al. Predicting the risk of pre-eclampsia between 11 and 13 weeks' gestation by combining maternal characteristics and serum analytes, PAPP-A and free beta-hCG. Prenat Diagn. 2010;30:1138–1142.
[28] Goetzinger K, Tuuli M, Cahill A, et al. Development and validation of a risk factor scoring system for first-trimester prediction of pre-eclampsia. Am J Perinatol. 2013;208:S300.
[29] Syngelaki A, Bredaki FE, Vaikousi E, et al. Body mass index at 11-13 weeks' gestation and pregnancy complications. Fetal Diagn Ther. 2011;30:250–265.
[30] Donders ART, van der Heijden GJMG, Stijnen T, et al. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59:1087–1091.
[31] Rubin D. Multiple imputation for non response in surveys. New York: John Wiley & Sons Inc.; 1987.
[32] Steyerberg EW, Borsboom GJJM, van Houwelingen HC, et al. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23:2567–2586.

[33] Janssen KJM, Vergouwe Y, Kalkman CJ, et al. A simple method to adjust clinical prediction models to local circumstances. Can J Anaesth. 2009;56:194–201.
[34] Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172:971–980.
[35] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128–138.
[36] Harrel F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; 2001.
[37] Bartsch E, Medcalf KE, Park AL, et al. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. BMJ. 2016;i1753.
[38] Rolnik DL, Wright D, Poon LC, et al. ASPRE trial: performance of screening for preterm pre-eclampsia. Ultrasound Obstet Gynecol. 2017;50(4):492–495.
[39] Oliveira N, Magder LS, Blitzer MG, et al. First-trimester prediction of pre-eclampsia: external validity of algorithms in a prospectively enrolled cohort. Ultrasound Obs. 2014;44:279–285.
[40] Farina A, Rapacchia G, Sterrantion AF, et al. Prospective evaluation of ultrasound and biochemical-based multivariable models for the prediction of late pre-eclampsia. Prenat Diagn. 2011;31:1147–1152.
[41] Herraiz I, Arbu J, Cama I, et al. Application of a first-trimester prediction model for pre-eclampsia based on uterine arteries and maternal history in high-risk pregnancies. Prenat Diagn. 2009;12:1123–1129.
[42] Rolnik DL, Wright D, Poon LC, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. N Engl J Med. 2017;377:613–622.