THE fullPIERS RISK PREDICTION MODEL FOR WOMEN WITH PRE-ECLAMPSIA: EXTERNAL VALIDATION, RECALIBRATION AND ADDED VALUE OF A NOVEL BIOMARKER (PLACENTAL GROWTH FACTOR)

by

Ugochinyere Vivian Ukah

MPH, The University of Sheffield, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Reproductive and Developmental Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2017

Abstract

The hypertensive disorders of pregnancy (HDPs), including pre-eclampsia, complicate up to 10% of pregnancies and are leading causes of maternal and perinatal morbidity and mortality. The fullPIERS model was developed to identify and quantify the risks of developing complications for women with pre-eclampsia in high-resource settings and to aid clinicians in managing such pregnancies. Prior to introducing the model into clinical practice, it is important to assess its external validity. Recalibration, if required, and addition of new biomarkers may be helpful to improve the predictive performance of the model.

The objectives of this thesis were (i) to assess the external validity of the fullPIERS risk prediction model for women with pre-eclampsia (ii) recalibrate the model if necessary, and (iii) to assess the incremental value of adding the biomarker, placental growth factor (PlGF), to the model.

Using abstracted medical records of women admitted into tertiary units in four high-income countries (HICs), the fullPIERS model was assessed for geographical and temporal validity. The model's predictive ability in women with a broader spectrum of disease including early-onset pre-eclampsia, other HDPs and low and middle-income countries was also assessed using existing cohorts. Good performance was interpreted based on discrimination (AUROCs $\geq$0.7) and calibration (slope $\geq$ 0.7). Stratification and classification accuracy of the model were also assessed.

The fullPIERS model showed good discriminatory performance on temporal and geographical validity (AUROCs >0.8) and also in broader HDPs (AUROCs >0.7). Medium to high likelihood ratios were estimated (>5 to >10) at a predicted probability cut-off of $\geq$30% for ruling in adverse maternal outcomes. Calibration was reduced in all cohorts (<0.7), except in the

temporal validation, suggesting a need for recalibration. Recalibration of the model improved the calibration performance but did not improve the discriminatory and stratification of the model. There was little incremental value of adding PlGF to the model.

The fullPIERS model has been successfully validated externally and can be implemented into routine clinical care in similar settings to inform clinical decisions. This will aid in appropriate allocation of care and resources to the patients, and contributing in reducing maternal morbidity and mortality resulting from pre-eclampsia.

Lay Summary

High blood pressure (hypertension) during pregnancy can cause severe complications, for example stroke or kidney failure for the mother, and fetal growth restriction or stillbirth for the baby. The ability to predict these adverse outcomes will ensure appropriate management and treatment of hypertensive pregnant women. This is by early identification of women with higher risks of complications arising from pregnancy hypertension and applying timely interventions. The goal of this thesis was to assess how valid a prediction model, the fullPIERS model, is and to improve the model predictive performance where possible, so that it can be implemented into practice to aid clinicians in managing such pregnancies. The results showed that the model is externally valid and will be useful for informing clinical decisions such as hospital transfer and timely delivery to prevent adverse outcomes. This may contribute to reducing the excess burden of maternal morbidity and mortality resulting from pre-eclampsia.

Preface

The fullPIERS validation study was approved by the UBC Clinical Research Board (**H07-02207)** and the research ethics boards of all partner institutions. All the data on which this dissertation is based were collected as part of the multicentre international fullPIERS (Pre-Eclampsia Integrated Estimate of RiSk) validation collaborative study. I contributed intellectually to the development of the study proposal and was responsible for primary study coordination including data transfer and sharing agreements and ethics application.

A portion of the introductory chapter was adapted from a previous book chapter publication of which I was the lead author and responsible for the screening of articles, writing and interpretation of the review: Ukah UV, Payne B, Côté AM, Hoodbhoy Z, von Dadelszen P. Risk factors and predictors of pre-eclampsia. In: Magee L, von Dadelszen P, Stones W., Mathai M, editors. The FIGO Textbook of Pregnancy Hypertension An evidence-based guide to monitoring, prevention and management. London: The Global Library of Women's Medicine; 2016, pp.75-100**.** L Magee and P von Dadelszen edited and approved the final version of the book and chapter.

One of the papers included in the Chapter 2 is published as: Ukah UV, Rocha BM, Mudenyanga C, Loquiha O, von Dadelszen P. The diagnostic performance of placental growth factor in women with suspected pre-eclampsia attending antenatal facilities in Maputo, Mozambique. I was also responsible for performing data analysis, initial interpretation of results and writing of the paper. One of the reviews in Chapter 2 has been accepted for publication as: U. Vivian Ukah, Jennifer A. Hutcheon, Beth Payne, Matthew D. Haslam, Manu Vatish, Mark Ansermino, Helen Brown, Laura A. Magee, Peter von Dadelszen. Placental Growth Factor as a Prognostic Tool in Women with Hypertensive disorders of Pregnancy: A Systematic Review. I

was the lead author, primarily responsible for conducting the literature search, data extraction and syntheses, and writing of the study.

A portion of chapter 5 has been published with me as the lead author: U. Vivian Ukah, Beth Payne, Tang Lee, Laura A. Magee, Peter von Dadelszen. External Validation of the fullPIERS Model for Predicting Adverse Maternal Outcomes in Pregnancy Hypertension in Low- and Middle-Income Countries. I was responsible for the data analyses, writing and initial interpretation of the study.

I was the lead investigator, under the direction of my thesis supervisor P von Dadelszen and committee members: J Hutcheon, L Magee and J M Ansermino, for all the work presented in this dissertation. As such, I contributed intellectually to the design, data collection, data analysis and initial interpretation of results and writing of all the chapters in this dissertation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACOG	American College of Obstetricians and Gynecologists
ASH	American Society of Hypertension
AOM	Association of Ontario Midwives
AST	Aspartate aminotransferase
ALT	Alanine aminotransferase
AUROC	Area Under the Receiver Operating Characteristic curve
BCW	British Columbia Women
CHIPS	Control of Hypertension in Pregnancy
CHWs	Community-based health workers
dBP	diastolic Blood Pressure
FINNPEC	Finnish Genetics of Preeclampsia Consortium
GA	Gestational Age
GCS	Glasgow Coma Score
GH	Gestational Hypertension
HELLP	Hemolysis Elevated Liver enzymes and Low Platelets
HDP	Hypertensive Disorders of Pregnancy
ISSHP	International Society for the Study of Hypertension in Pregnancy
IDI	Integrated Discrimination Improvement
IV	Intravenous
JRH	John Radcliffe Hospital
LDH	Lactate dehydrogenase

LMIC	Low- or Middle- Income Country
LR-/+	Positive or negative likelihood ratio
MICE	Multiple Imputation by Chained Equations
NGAL	Neutrophil gelatinase-associated lipocalin
NHBPEP	National High Blood Pressure Education Program
NICE	National Institute for health and Clinical Excellence, UK
NICU	Neonatal Intensive Care Unit
NPV	Negative Predictive Value
NRI	Net Reclassification Index
OR	Odds Ratio
PELICAN	Pre-EcLampsIa: Clinical ApplicatioN
PETRA	Pre-Eclampsia Triage by Rapid Assay (for Alere-PETRA)
PETRA	Pre-eclampsia Trial Amsterdam (for Dutch PETRA)
PIERS	Pre-eclampsia Integrated Estimate of RiSk
PPV	Positive Predictive Value
PREP	Prediction of complications in early-onset pre-eclampsia
PRE-EMPT	Pre-eclampsia-Eclampsia Management, Prevention and Treatment
RCT	Randomized Controlled Trial
RR	Relative risk Ratio
SOGC	Society of Obstetrics and Gynecologists Canada
SOMANZ	Society of Obstetric Medicine of Australia and New Zealand
sBP	Systolic Blood Pressure
SGA	Small-for-gestational age

Spot	Pr/CR Spot urinary protein to creatinine ratio test
SpO_2	Blood oxygen concentration by pulse oximetry
Sens	Sensitivity
Spec	Specificity
VIPER	Vancouver Interdisciplinary Pre-eclampsia, Eclampsia and growth Restriction working group
WHO	World Health Organization

Acknowledgements

Several people have contributed immensely to the work presented in this thesis. Firstly, I would like to thank my supervisor, Dr. Peter von Dadelszen and my thesis committee members, Drs. Jennifer Hutcheon, Laura Magee and Mark Ansermino, who have been my able teachers and guided me throughout the process. The project would not have been conceived or been a success without their continuous support, direction, and mentorship.

I would also like to thank all the collaborators (listed in **Appendix A**), who contributed to the datasets used in this study and supported the project intellectually and I am grateful to the research assistants and midwives who were involved in data collection for this thesis. It was a wonderful experience working with this large and diverse group from all over the world. I am extremely grateful to Dr. Beth Payne for her mentorship and the rest of the VIPER-PRE-EMPT group for their continuous support.

I feel blessed and thankful for the four year fellowship from the University of British Columbia and for the Canadian Institutes of Health Research (CIHR) operating grant which funded this project. Very importantly, I must acknowledge the women whose data contributed to this project as this thesis would not have been possible without them, as well as the fullPIERS working group that developed the model from which this thesis is built upon.

Finally, I would like to thank my unconditional support system – my mother, Mrs. Bridget Ukachi Ukah, and my brothers and sisters – Iheanyi, Ada, Edozie, Chimdindu and Odichinma (my personal informant); their love and prayers kept me going throughout.

Dedication

This thesis is dedicated to my father Engr. Chinenyem Boniface Ukah; your love and memories will forever remain in our hearts. To God be the glory forever, Amen.

A father of six, a figure for many

Classified with many great words

But could not be qualified by a single one

Valour and great brains, some described him

Witty and hardworking, others would add

To each child, he gave different gifts

And with each name, he shared their roles

To the eagle that God gave him, He helped to fly

Beyond the skies would be her starting point

He went to rest but not before his tasks were done

A legacy he made sure to leave

A soul that stayed calm with the Lord

With each feather she grew, he beamed with pride

How she knows of this – his voice in her heart.

Chapter 1: Introduction

1.1 Classification of the hypertensive disorders of pregnancy

Hypertension is one of the most commonly encountered problems in pregnancy and is generally defined as high blood pressure - systolic blood pressure (SBP) ≥ 140mmHg and/or a diastolic pressure (DBP) ≥ 90mmHg, measured twice at least 4 hours apart.[1,2] The hypertensive disorders of pregnancy (HDPs) are broadly classified into chronic or pre-existing hypertension, gestational hypertension, pre-eclampsia, and superimposed pre-eclampsia.[1,3] Chronic hypertension is defined as having high blood pressure diagnosed before 20 weeks of gestation and gestational hypertension is defined as having high blood pressure diagnosed after 19^{+6} weeks (i.e. ≥ 20^{+0} weeks) of gestation.[4,5] Pre-eclampsia is classically defined as new hypertension arising from ≥ 20^{+0} weeks of gestation with proteinuria and superimposed pre-eclampsia is classically defined as pre-existing hypertension with accelerated hypertension or with new or accelerated proteinuria.[1,6]

A systematic review of 13 international and national clinical guidelines reported that the definitions of chronic and gestational hypertension were consistent across the guidelines; however the definition of pre-eclampsia and its severity lacked consensus.[4] Some of the guidelines that were reviewed included the American College of Obstetricians and Gynecologists (ACOG),[7] Association of Ontario Midwives (AOM),[8] National Institute for Health and Clinical Excellence (NICE),[9] Society of Obstetricians and Gynaecologists of Canada (SOGC),[1] and World Health Organization (WHO).[6] For the definition of pre-eclampsia, some guidelines also include gestational hypertension with one or more adverse features such as hyperuricaemia, feto-placental abnormalities and maternal symptoms associated with pre-eclampsia;[1,7] in addition, the threshold of proteinuria for the diagnosis of pre-eclampsia was not consistent. Similarly, the

severity criteria for pre-eclampsia also differ across guidelines; some guidelines use severe hypertension, defined as BP ≥160/110 mmHg and other biochemical impairments (e.g. NICE, ACOG), or include end-organ complications such as eclampsia (e.g. SOGC), while heavy proteinuria has been used to define severity in some guidelines (e.g. WHO and AOM). In the AOM guideline, pre-eclampsia occurring before 34 weeks (early-onset pre-eclampsia) is also a severity criterion.[8]

These differences in the diagnosis of pre-eclampsia reflect the lack of clarity of the pathogenesis of the disease, which is discussed in the following section. The definition of pre-eclampsia used in the primary study on which this thesis is based on was primarily the SOGC guidelines.

1.2 Global burden of HDPs

HDPs complicate about 10% of pregnancies globally and are among the top three causes of maternal and neonatal morbidities and mortality worldwide **Figure 1-1**.[2,10] Approximately 6% of pregnancies are reported to be complicated by gestational hypertension while ~1% is complicated by chronic hypertension and 2% by pre-eclampsia.[1] Pre-eclampsia is the most severe form of HDP and accounts for about 50,000 maternal deaths annually.[2] Although the majority of the maternal mortality resulting from the HDP occurs in low- and -middle-income countries,[11] there are still associated poor outcomes of pregnancy reported in high-income countries.[10] In fact, HDPs are reported to be the cause of 16.1% of maternal deaths in developed countries and the leading cause of planned pre-term delivery in Canada.[12] Thus, pre-eclampsia and other HDPs pose a major challenge globally.

1.2.1 Complications

The complications that arise from pre-eclampsia and other HDP can result in life-endangering or ending consequences for both the mother and the fetus if not managed efficiently.[3] The adverse maternal complications from HDP include eclampsia, stroke, acute respiratory distress syndrome (ARDS), cerebral haemorrhage and infarction, pulmonary oedema, renal dysfunction, hepatic haematoma and placental abruption.[1,6] Adverse fetal and neonatal outcomes include intrauterine growth restriction (IUGR), stillbirth, low birth weight or small-for-gestational age (SGA), neonatal death, and hypoxic-ischaemic encephalopathy. Preterm delivery is also highly associated with pre-eclampsia; about 10% of all pre-term births result from HDPs and these could lead to later complications as a result of prematurity such as cerebral palsy and chronic lung diseases.[13]

Due to the associated high morbidity and mortality, the HDP are of great health concern and constitute a burden not only to clinicians but also to women, families and societies globally.

1.3 Pathophysiology

The pathogenesis of pre-eclampsia is unclear; however, it is understood to be a maternal syndrome resulting from a mismatch between fetal demands and utero-placental supply and can affect multiple organs.[2,13,14] Pre-eclampsia can originate from 'placental' or 'maternal' factors or a combination of both. 'Placental pre-eclampsia' or early-onset pre-eclampsia (occurring <34 weeks gestation) arises from poor placentation in the early stages of pregnancy.[13] Poor placentation occurs when there is inhibition of invasion and remodelling of the maternal spiral arteries by the extravillous cytotrophablast (found in the placental tissue).[14] This invasion and remodelling of the maternal vessels is important to increase blood supply to the fetus via the placenta. Thus, inadequate placentation leads to an imbalance in the uteroplacental blood flow,

which results in oxidative stress, maternal endothelial damage, and multi-organ dysfunctions. Maternal pre-eclampsia or late onset pre-eclampsia is associated with maternal predisposition to arterial disease resulting in a hyper-inflammatory state during pregnancy.[13]

1.4 Risk factors and predictors of HDPs

The risk factors for pre-eclampsia and other HDPs include primiparity, obesity, low sperm exposure, multiple gestation, extreme maternal age (≤19 or ≥35 years) and family or previous history of HDPs and pre-existing conditions such as pre-gestational diabetes mellitus and renal disease.1,15 In a review of predictors for pre-eclampsia, no single marker could effectively predict the development of pre-eclampsia.15 The most promising predictor was Doppler ultrasound and combinations of clinical markers with angiogenic factors in multivariable models.

1.5 Management of the HDP

The management of pre-eclampsia includes treatment of symptoms using antihypertensive and anticonvulsant therapies (i.e. expectant management) but delivery is the only cure.[2,3] Symptomatic management of HDPs during expectant management does not necessarily prevent poor outcomes for the mother, as it often does not change the progression of the disease.[2] Although a few studies have reported that treatment of high blood pressure during pregnancy may reduce the risk of severe hypertension for women with chronic and gestational hypertension,[16,17] there is a lack of evidence for the prevention of other adverse maternal outcomes.[18] On the other hand, inducing delivery which occurs in interventionist management may lead to premature delivery which is associated with increased health risks for the baby, especially when remote from term. When pre-eclampsia occurs in earlier gestational ages (<32 to 34 weeks), the decision to induce labour or delivery (interventionist) or carry out expectant management is usually dependent on the subjective clinical impression of the disease severity

and personal experience/expertise of the clinician.[19] Between 34 to 37 weeks of gestation, the length of time to expectantly manage the woman is also not clear; however, the HYPITAT-II study has shown that adverse perinatal events are reduced by a policy of expectant management even near term. [20] This is due to insufficient clinical evidence of the clear benefits or disadvantages of expectant management versus interventionist approach to the mother or the fetus. These indicate the need for more evidence-based information to guide decisions about the management of the pre-eclampsia at such gestational ages.

1.6 Clinical prediction models

Prediction models are becoming increasingly common in medicine.[21] The use of prediction models may be more beneficial in risk prediction than using individual factors, as they have the advantage of combining various factors to potentially provide more accurate and personalized predictions. Prediction models are especially useful for the interpretation of risks and communication between clinicians and patients as health care is moving towards patient-centred care.[21,22] Before any developed model is implemented into practice, it needs to be both internally and externally valid to ensure that the predictions of outcomes are accurate enough to inform management.[22,23]

1.6.1 Prediction of complications from HDPs

The ability to predict adverse outcomes resulting from HDPs could guide management decisions, such as the need for transfer to a higher level of care (e.g. from a primary or secondary centre (Levels I and II maternity units) to a tertiary unit (Level III)), and timing of delivery.[19] A few studies have attempted to predict adverse outcomes resulting from HDPs using individual risk factors or other biomarkers. These studies are discussed in **Chapter 2**; most of these single

markers were insufficient to predict adverse maternal outcomes from HDPs. One promising multivariable prognostic models that has been developed to aid in the care of women with hypertensive disorders of pregnancy is the fullPIERS model.[24]

1.6.1.1 The fullPIERS (PIERS – Pre-eclampsia Integrated Estimate of RiSk) model

The fullPIERS model was developed to predict adverse maternal outcomes occurring within 48 hours of a woman's admission to a tertiary hospital for pre-eclampsia. The rationale behind this PIERS project was that correctly identifying individual women's risk of complications before they happen would improve the clinician's ability to counsel patients on timing of delivery and use of other interventions in order to avoid those complications. The multivariable model was developed using a cohort of 2,023 women, admitted to tertiary level facility with pre-eclampsia in high income countries (HICs): Australia, Canada, New Zealand and the United Kingdom (UK).

The study's primary adverse maternal outcome was a composite of severe maternal complications which were agreed upon during a Delphi consensus for the model development study. The outcomes included central nervous system (CNS), hepatic, renal, cardiovascular and respiratory outcomes; a full list of the outcomes is online (https://pre-empt.cfri.ca/monitoring/fullpiers) and in the **Appendix B**. Further details of the fullPIERS cohort have been described in **Chapter 3**.

Six predictor variables were included in the model: gestational age at admission for pre-eclampsia, chest pain or dyspnoea, oxygen saturation (SpO_2), platelet count, serum creatinine, and serum aspartate transaminase (AST). The fullPIERS model was internally validated and had excellent discriminatory performance, with an apparent area under the receiver-operating

characteristic curve (AUROC) of 0.88 (95% confidence interval [CI]: 0.84–0.92) and minimal estimated optimism of 0.02 upon internal validation using Efron's bootstrap method. The fullPIERS model equation is shown in **Box 1.**

The fullPIERS model also fulfilled the Altman and Lyman criteria used to assess the purpose of predictive models which include: (1) the need to guide clinical decision making (including management and counselling of women with pre-eclampsia), (2) improve the understanding of the disease process (by allowing basic scientists to categorise women with pre-eclampsia into meaningful subgroups associated with differential maternal risk), (3) define at-risk groups based on prognosis (i.e., risk stratification) and (4) predict outcomes in pre-eclampsia pregnancies.

These statistically-robust methods behind the model development and its good predictive performance make the fullPIERS model a promising tool for predicting adverse maternal outcomes in women with pre-eclampsia. However, the model needs to be externally validated before it can be used in clinical practice, that is, its performance should be assessed in a different population other than the cohort in which it was developed (i.e. the development cohort). Validation ascertains that the model truly predicts adverse maternal outcomes in women admitted with pre-eclampsia.

Box 1: The fullPIERS Logistic Regression Equation: 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}; platelets²)+(1·01×10–²; aspartate transaminase)+(–3·05×10–^{6}; AST²)+(2·50×10–^{4}; creatinine×platelet)+(–6·99×10–^{5}; platelet×aspartate transaminase)+ (–2·56×10–³; platelet×SpO_2)

1.7 Thesis objectives

Several prediction models are being developed although many are not implemented into clinical practice due to poor development techniques, poor performance, and lack of internal and external validity.[21,25]

The overall goal of the research presented in this thesis was to assess the external validity of the previously developed fullPIERS risk prediction model for women with pre-eclampsia and to improve the model where possible, so as to enable its implementation into clinical practice for guiding the management of women with pre-eclampsia. One of the ways tested in this thesis to improve the model was the addition of a new biomarker, the placental growth factor (PlGF); recalibration of the model was also carried out.

Thus, the specific objectives for this study were:

1. To assess the predictive ability of the fullPIERS model in a different cohort than the one in which it was developed (external validation)

2. To assess the model in women presenting with pre-eclampsia in the same centre where it was developed, from a more recent time period (temporal validation).

3. To recalibrate the model where necessary (updating of intercept, and slope if required)

4. To determine if the predictive ability of the fullPIERS model can be increased by incorporating a new biomarker, the Placental Growth Factor (PlGF)

The proposed hypotheses were as follows:

1. The fullPIERS model will perform well (AUROC ≥ 0.70) in identifying women at relatively high risks of adverse maternal outcomes from pre-eclampsia across different settings.

2. Additionally, PlGF will add a significant incremental value to the performance of the fullPIERS model.

These objectives are met through several stages of the model assessment in the subsequent chapters. In Chapter 2, a systematic review is presented on the prognostic tests for adverse maternal outcomes for women with pre-eclampsia and another review is presented on the use of PlGF as a prognostic test for women with HDPs. These reviews were conducted to summarize the results from available prognostic tests for HDPs as well as to justify the need for the external assessment of the fullPIERS model and the addition of PlGF to the model in this thesis. In Chapter 3, the datasets and the model performance evaluation methods used throughout this thesis are described. In Chapter 4, temporal and external (geographical) validity assessments of the model are carried out and the results are presented. In Chapter 5, the model is assessed in a broader context including a sub-group of pre-eclampsia and other types of HDPs (domain validation) and Chapter 6 presents the recalibration and addition of PlGF to the model. Finally in Chapter 7, possible reasons for the observed model performance are explored and discussed.

1.8 Summary

The HDPs, especially pre-eclampsia, constitute a significant burden to maternal health globally. Adequate management and timely delivery might aid in averting adverse outcomes resulting from pre-eclampsia and other HDPs. The ability to predict which women, diagnosed with HDPs, are at higher risks of experiencing complications could improve management of the disease. The fullPIERS model appears promising but requires external validation before implementation. If externally valid, the model can be implemented in clinical settings to guide healthcare providers. To increase utilization after successful validation, the model can be presented in simpler formats such as decision trees and calculators on mobile application devices, along with suggested

actions for risk stratifications.[18] We hope that the fullPIERS model aid in reducing the global burden of adverse maternal and perinatal events associated with the HDPs.

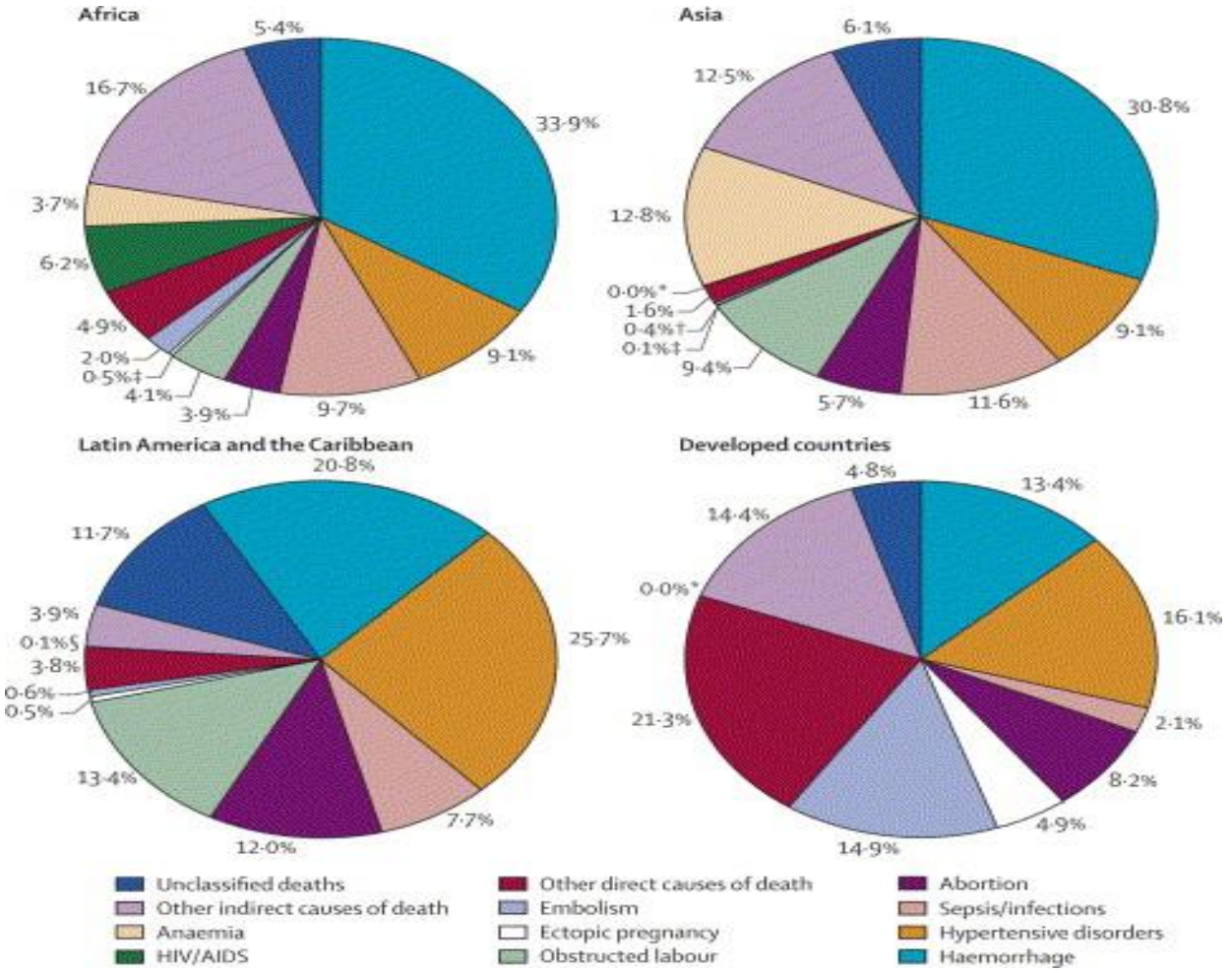


Figure 1-1 Geographical variation in distribution of causes of maternal deaths[1]

[1] Adapted from WHO analysis of causes of maternal death: a systematic review by Khan et. al.12

Chapter 2: Prognosis of the hypertensive disorders of pregnancy

2.1 Background

Predicting the onset of complications of HDPs could aid in timely interventions such as increased surveillance, treatment of symptoms, transfer to higher care facility and delivery when necessary, which in turn could reduce morbidity and mortality from the HDPs.[19,24] Maternal risk factors used as criteria for severity classification by some international clinical practice guidelines do not accurately identify women at high risk of developing maternal complications.[26] The tests reported in prediction studies range from single markers to multiple markers combined in multivariable prediction models. Regardless of the prediction method used, there is a need for the results from these studies to be summarized and compared. This will determine if they give meaningful and accurate information to assist clinicians in the management of the HDPs.

In **Section 2.2** of this chapter, a review of tests for the prediction of solely adverse maternal outcomes resulting from HDPs is presented while in **Section 2.3**, a review of PlGF for the prediction of both maternal and fetal adverse outcomes resulting from HDPs is presented.

2.2 Prediction of adverse maternal outcomes from HDPs

Several systematic reviews have assessed the predictive ability of individual variables such as uric acid, maternal symptoms, and liver function tests for maternal and fetal complications resulting specifically from pre-eclampsia.[27-29] However, reviews assessing predictors for maternal complications resulting from all types of HDPs are unknown. This broader disease inclusion is important, as other HDPs also contribute substantially to the burden of the disease. In addition, these reviews were conducted between 2006 and 2011 and since then, the classification system for HDPs, particularly pre-eclampsia, has evolved. Furthermore, the studies included in these reviews solely assessed potential univariable predictors, thus the need to also

review potential predictors combined in multivariable models. Therefore, a systematic review of studies reporting the predictive ability, for both single and combined markers, of adverse maternal outcomes in women with HDPs was conducted.

2.2.1 Methods

2.2.1.1 Eligibility criteria

The studies of interest were conducted on women with a HDP: pre-eclampsia, gestational hypertension, or chronic (pre-existing) hypertension, as defined by the study (with study definitions documented). The predictors of interest were any tests measured to predict adverse maternal outcomes from HDP. The adverse maternal outcomes considered were severe complications from the HDPs, which had been agreed upon in a Delphi Consensus in the PIERS study.[22] In addition, postpartum haemorrhage (PPH) and disseminated intravascular coagulation (DIC) were considered as these outcomes have been subsequently reported to be strongly linked with HDPs.[30] Detailed inclusion and exclusion criteria and full list of outcomes of interest are shown in **Table 2-1**.

2.2.1.2 Search and selection strategy

Databases searched were MEDLINE, Embase, CINAHL, EBM and Cochrane Library databases from their inception to December 2016. Google Scholar and grey literature sources (such as University of British Columbia circle, government websites, etc.) were also searched for other potential articles. Web of Science was used for citation tracking of review articles and eligible articles, to capture any articles that were not identified through the electronic search. The search terms included both relevant subject heading terms and key words related to the HDPs, with methodological filters to identify prognostic test studies for maternal complications (see **Appendix C**).

All retrieved articles were screened independently for eligibility by two reviewers (UVU - myself and DAD), first by title and abstract and then, by reviewing the full articles. Final selections were compared and any conflicts resolved by discussion and/or by a third reviewer (BP).

The predictive measures used were sensitivity, specificity, likelihood ratios (LRs), and AUROCs.[31] Negative predictive values (NPVs) and positive predictive values (PPVs) were not evaluated in this review because of their dependence on outcome prevalence. Studies that reported none of these predictive measures were included only if adequate data were provided to calculate these measures. Studies reporting both maternal and fetal outcomes as a combined outcome were excluded except in cases where the test prediction performance for the maternal outcomes could be separated. Studies that included any of the HDPs as one of the outcomes were also excluded.

2.2.1.3 Data extraction and assessment of study quality

For each eligible study, information on population characteristics, tests used as predictors, measures and accuracy of prediction were extracted. Methodological quality assessment of the included studies was carried out using the QUIPS (Quality in Prognostic Studies) tool,[32] which has been validated and used in similar studies.[33] The relevant study aspects that were scrutinized included methods of sampling and recruitment, adequate description of tests and outcomes, complete follow-up or handling of missing data explained, and sample size. In total, there were eight questions considered and one point was awarded for each assessment question that was met. In addition, studies reporting multivariable prediction models were assessed for internal validation. Studies with a total score of ≥ 7 were considered as having a low risk of bias, 4-6 as medium risk of bias, and <4 as high risk of bias.

2.2.1.4 Data synthesis

Two by two tables for each included study were constructed, cross-classifying test results and the occurrence of adverse maternal outcomes. Predictive measures were either retrieved directly from the studies or constructed from raw data and 2 × 2 tables. LRs were used to provide interpretations for clinical usefulness as a measure that is independent of disease prevalence; for positive LRs (LR+), LRs of 5-10 and >10 were interpreted as having moderate and strong evidence to 'rule in' the disease respectively while for negative LRs (LR-), LRs of 0.1-0.2 and <0.1 were interpreted as having moderate and strong evidence to 'rule out' the disease respectively.[34] An AUROC ≥0.70 was considered to reflect good discriminatory ability for multivariable models.[35] Wherever possible, meta-analyses were conducted for similar tests predicting similar outcomes and having 3 or more 2 × 2 tables.[36,37]

2.2.2 Results

2.2.2.1 Literature Search and identification results

Of 2137 articles retrieved, 28 primary articles were included in the review. Reasons for excluding studies were: inclusion of a HDP e.g. HELLP syndrome as an outcome in the study or having inclusion criteria that were not restricted to women with a HDP (N=6), or did not present maternal and fetal outcomes separately (combined only) (N=12), or did not present data in a 2 × 2 table to calculate the diagnostic tests characteristics of interest (N=3) **(Figure 2-1)**.

2.2.2.2 Characteristics of included studies

The included articles were published between 1988 and 2016 (**Table 2-2)**. Nine were multicentred studies and 19 studies were conducted in single centres. Most studies (26/28, 92.9%) used a cohort design, of which 21/28 (75%) were prospective. There was one randomized trial and case-control study each. Data were collected from a range of countries

including Australia (N=8), United Kingdom (N=8), Canada (N=7), New Zealand (N=7), USA (N=6), South Africa (N=4), India (N=2), Netherlands (N=1), Pakistan (N=2), and one each from Iran, Spain, Tunisia, Turkey, Uganda, Mexico, and Brazil. All included studies were published in English except for one study that was in French. Overall, the number of independent women in the included studies was 7,562, with a mean or median gestational age at admission or recruitment ranging from 23 weeks to 36 weeks. The mean maternal age ranged from 23 to 35 years old, and 13% to 89% were nulliparous.

2.2.2.3 Definitions of HDPs

Eight studies included women diagnosed with either chronic hypertension (N=4) or gestational hypertension (N=4), while the remaining studies reported solely on women with pre-eclampsia (including HELLP syndrome) and/or superimposed pre-eclampsia. The definition of pre-eclampsia varied by the reference guideline used: the International Society for the Study of Hypertension in Pregnancy (ISSHP)[5] (N=9), SOGC[1] (N=9), ACOG[7] (N=8), NICE[9] (N=1), or National High Blood Pressure Education Program (NHBPEP)[38] (N=1) guidelines.

2.2.2.4 Quality of studies

The studies scored well with respect to adequacy of population selection description, appropriateness of the patient spectrum/representativeness, and adequacy of test and outcome descriptions **(Figure 2-2)**. However, of the 28 studies, only 13 mentioned complete follow-up or explained withdrawals, 10 reported on handling of missing data, six reported sample size calculations, and two of six multivariable model studies reported internal validation; there were also attempts at external validation for these two models. As a result, only eight studies were ranked as having at low risk of bias, 18 at medium risk, and two as high risk.

2.2.2.5 Outcomes and data synthesis

The prevalence of adverse maternal outcomes in the studies ranged from 1.1% to 34.2%. Nine studies reported on single outcomes, most commonly eclampsia (N=6), and placental abruption (N=6). Most studies (19/28, 67.9%) reported on composite outcomes; these usually included the common single outcomes as well as thrombocytopenia, PPH, ascites, and hepatic rupture.

2.2.3 Univariable predictors

2.2.3.1 Blood pressure

One of the signs evaluated was blood pressure, which was assessed in three studies[37-41] as systolic (N=2), diastolic (N=1), or mean arterial pressure (MAP, N=1). The outcomes being predicted in these studies were eclampsia and placental abruption, for women with either pre-eclampsia or mild chronic hypertension. The cut-off for SBP evaluated were >140 and ≥ 160 mmHg while the cut-off for DBP was > 90 mmHg; MAP was assessed at > 105 mmHg. Although significant associations (*p*-values <0.05) between blood pressure and adverse outcomes were presented in these studies, none of them showed a clinically useful measure for blood pressure as a prognostic test for adverse maternal outcomes (**Table 2-3)**.

2.2.3.2 Proteinuria

Proteinuria was assessed in six studies,[42-45] using measurements of 24h urinary protein excretion (N=5), spot protein/creatinine ratio (N=1), spot albumin/creatinine ratio (N=1) and/or urinary dipstick testing (N=3). Only the study by Bouzari *et al*[43] reported a moderate LR- for ruling out placental abruption using 24-hour urine proteinuria, at a cut-off of 1750mg [LR- of 0.1 (95% CI: 0.0–0.6)] with an AUROC of 0.78. No other study reported a clinically useful measure for ruling in or out adverse maternal outcomes using proteinuria testing. The findings presented in this thesis are similar to the study by Thangaratinam *et al*[27], which reported that proteinuria was a

poor predictor of maternal complications in pre-eclampsia based on the pooled positive and negative LRs in their study. In concordance, another review by Morris *et* al[47] stated that there was insufficient evidence to recommend proteinuria as a prognostic test for the prognosis of adverse maternal outcomes in pre-eclampsia. Proteinuria is also not recommended as a test for the prediction of adverse maternal outcomes for women with HDPs by ACOG, and SOGC guidelines.

2.2.3.3 Other signs and symptoms

Maternal symptoms evaluated were: headache (N=3 studies),[39,48,49] visual disturbance (N=3),[39,48,49] nausea or vomiting (N=2),[48,49] right upper quadrant pain or epigastric pain (N=2), chest pain or dyspnoea (N=2),[48,50] abdominal pain and vaginal bleeding (N=1),[48] and hyperreflexia (stated as "vivid" deep tendon reflexes in the study)[39] or "non-specific viral symptoms"[49] (which was not defined in the study) (N=1 study each). The other sign evaluated as a univariable test was oxygen saturation in one study.[50] Only non-specific viral symptoms had moderate LR+ for ruling in composite adverse maternal outcomes while headache, visual symptoms, and hyperreflexia each had moderate LRs (-) for ruling out eclampsia (LRs between 0.1 and 0.2). Non-specific viral symptoms and oxygen saturation of <93% also had reported a AUROCs of ≥0.7 suggesting good discriminatory ability for the prediction of composite adverse maternal outcomes. The usefulness of each of these symptoms was demonstrated in only one study.

In a systematic review of maternal symptoms as predictors of adverse outcomes, epigastric pain and visual disturbance were reported to be the most useful predictors based on their AUROCs; however, this was not the case in this review in this thesis. In that review by Thangaratinam *et al*,[51] HELLP syndrome was considered as an adverse outcome and was one of

the most reported outcomes in the review whereas HELLP syndrome was one of the inclusion criteria for HDPs in this thesis review, because it has been recognised as part of the spectrum of pre-eclampsia rather than an outcome. Therefore, majority of the studies from the review by Thangaratinam *et al*[51] were excluded in this review and it is possible that the performance of epigastric pain in their systematic review was related to women with HELLP rather than predictive of adverse maternal outcomes that measure end-organ failure.

2.2.3.4 Laboratory tests

The laboratory tests assessed included: platelet count (N=4 studies),[41,49,52,53] serum creatinine (N=1), serum uric acid (N=3), international normalized ratio (INR, N=1), aspartate transaminase (AST, N=4), alanine transaminase (ALT, N=2), lactate dehydrogenase (LDH, N=2), serum albumin and total bilirubin (N=1 each). None of the laboratory tests had a useful LR+ to rule in adverse maternal outcomes. Only serum uric acid had a moderate LR- for ruling out eclampsia in one study (LR- of 0.1 (95% CIs: 0–0.9). Contrary to the findings in the review by Koopmans *et al*,[29] uric acid did not show any clinical usefulness in the prediction of maternal outcomes in our review except for eclampsia. Similar to the previous systematic review on maternal symptoms,[51] HELLP syndrome, which was an inclusion criterion in this review, constituted a majority of the adverse outcomes in these previous reviews. However, our findings are similar to the review by Thangaratinam *et al*,[54] which also reported poor pooled LR (LR+ of 2.1 and LR- of 0.38) for predicting adverse maternal outcomes using uric acid.

AST, and ALT, and LDH were reported to have good discriminatory abilities, with AUROCs of ≥0.70 for prediction of adverse maternal outcome in only one study by Kozic *et al*.[55] A systematic review by Thangaratinam *et al*[28] reported that liver enzyme tests (AST, ALT, and LDH) were moderate predictors of combined maternal and fetal complications in women

with pre-eclampsia. Although AST, ALT, and LDH did not have any strong clinical utility in univariable analyses based on LRs, their AUROCs in the individual studies suggest consideration in future studies. This is also in line with the NICE guideline, [12] which recommends that more studies for kidney and liver function, and coagulation for the prediction of adverse outcomes are needed.

2.2.3.5 Biomarkers

Neutrophil gelatinase-associated lipocalin (NGAL) was evaluated as a predictor in one study for the prediction of composite adverse maternal outcomes.[56] This biomarker did not show any clinically useful measures to either rule in or rule out adverse maternal outcomes. The other biomarkers evaluated were PlGF independently (N=1 study)[57] or combined with soluble fms-like tyrosine kinase-1 (sFlt1) as a ratio (N=3 studies).[58-61] The prognostic value of PlGF is discussed in more detail in **Section 2.3**.

2.2.4 Multivariable predictors

Six studies evaluated a combination of multiple variables to predict a composite of adverse maternal outcomes. Four of these multivariable studies were part of the PIERS studies: the fullPIERS model,[24] miniPIERS model,[62] an extended miniPIERS model with SpO_2,[63] and a combined cardiorespiratory symptom model by Millman *et al*[50] for the prediction of the PIERS composite outcome; the outcomes in two other studies by Chan *et al*[64] and Girling *et al*[65] also included some components of the PIERS outcomes such as renal failure, thrombocytopenia, liver disease and pulmonary oedema. The miniPIERS model and the extended miniPIERS model with SpO_2, were the only multivariable models that included women with all HDPs, while the others included women with (superimposed) pre-eclampsia. The most commonly included predictors in these models were chest pain and dyspnoea (N=4 models), oxygen saturation and gestational age

(N=3), and AST (N=2). Three of the multivariable models (fullPIERS model, miniPIERS model, and extended miniPIERS model with SpO_2) reported moderate to high LR+ for ruling in adverse maternal outcomes (LR+ of 5 and above) and four of them (models by Millman, fullPIERS model, miniPIERS model, extended miniPIERS model with SpO_2) reported AUROCs of ≥0.7; all of these were predicting PIERS adverse outcomes. The fullPIERS model by von Dadelszen *et al*[24] had the highest AUROC of 0.88 (95% CI: 0.84–0.92) and also had the highest LR+ of 26.5 to strongly rule in composite adverse maternal outcomes at a predicted probability of ≥ 30%.

2.2.5 Discussion

2.2.5.1 Main findings

Overall, any univariable test with moderate prognostic performance was either assessed in only one study (e.g. hyperreflexia and "non-specific viral symptoms"), or demonstrated usefulness in only one study of multiple. As such, individual tests were interpreted as lacking strong evidence of clinical usefulness. However, the most promising tests for consideration in multivariable models were oxygen saturation, headache, visual symptoms, AST, chest pain or dyspnea, and gestational age, based on their inclusion in well performing multivariable models. Multivariable models such as the miniPIERS and the fullPIERS model showed the most promising prognosis for maternal outcomes.

2.2.5.2 Strengths and limitations

This review summarized the performance of potential predictors of maternal complications among women with all types of HDPs using updated definitions for HDPs and with no restrictions on language or year of publication. Including all HDPs in the review improved the clinical applicability because it includes a broader population of women at risk and not all

women initially present with pre-eclampsia at admission. Multivariable models which have not been assessed in any previous reviews were also reviewed.

The majority of the included studies in this review were deemed to be of low or moderate quality. Many of the studies were underpowered and some of the multivariable models were not externally validated; thus, the results from these studies may not be applicable in a different setting or population. However, these studies were not excluded because of the interest in any tests with potential maternal prognostic value for HDPs and also due to sparse literature in the study area. The only articles not included were the studies that did not meet the inclusion criteria for severe maternal outcomes; for example, a prognostic study by Allotey *et al*,[66] that included preterm delivery as one of their composite maternal outcomes, was excluded from the review because preterm delivery was not included in our pre-specified list of adverse maternal outcomes. The methodological issues in the included studies and the heterogeneity in population characteristics affect the ability to draw any strong conclusions in the review.

2.3 Placental Growth Factor for prognosis in HDP

Several studies have established that the placenta plays an essential role in the development of HDPs, especially pre-eclampsia.[67,68] PlGF has been shown to be an influential component of prediction of pre-eclampsia at 11-13 weeks.[69] Many studies have reported that there is an angiogenic imbalance in pregnancies complicated by pre-eclampsia and/or intra-uterine growth restriction (IUGR).[70-73] In such pregnancies, concentrations of proangiogenic factors such as the placental growth factor (PlGF) and vascular endothelial growth factor (VEGF) are decreased in maternal circulation while antiangiogenic factors such as soluble fms-like tyrosine kinase-1 (sFlt1), also known as VEGFR-1, and soluble endoglin (sENG) are increased. These findings

have led to speculations that angiogenic factors might be useful in the prediction of pre-eclampsia and related adverse outcomes.[2,67]

The majority of studies of PlGF testing have focused on either prediction of pre-eclampsia or confirmation of the diagnosis once pre-eclampsia is suspected. A systematic review suggested that incorporation of this biomarker into a clinical multivariable model may improve prediction of pre-eclampsia.[74] Also, PlGF is cost-saving if used before 35 weeks' gestation for predicting pre-eclampsia requiring delivery within a specified time.[75] Fewer studies have reported on the use of PlGF for the prediction of adverse outcomes among women already diagnosed with pre-eclampsia or other HDPs.[61,76] A review on the use of PlGF as a prognostic tool in HDPs is needed to summarize the results from these prognostic studies in order to establish the clinical utility of PlGF in the identification of women with HDPs at highest risk of experiencing adverse outcomes. Therefore, a systematic review of the findings from studies reporting the use of PlGF as a prognostic test for women with suspected or confirmed pre-eclampsia was conducted.

2.3.1 Methods

2.3.1.1 Eligibility criteria

Any study using PlGF (either as an independent marker or combined with other angiogenic or clinical markers) as a prognostic test for adverse health outcomes (maternal or/and fetal outcomes) in women with suspected or confirmed HDPs and reporting predictive performance measures (as described in **Section 2.2**), or providing sufficient data to calculate these measures, was considered eligible. The outcomes of interest were severe maternal outcomes, similar to the review above as well as fetal outcomes related to HDPs. Detailed inclusion and exclusion criteria and full list of outcomes of interest are shown in **Table 2-4**.

2.3.1.2 Search and selection Strategy

The same databases and sources as mentioned earlier in **Section 2.2** were searched from their inception to January 2017 to identify eligible articles. The search terms included both MeSH terms and key words related to the HDPs, with methodological filters to identify prognostic test studies for maternal complications (**Appendix C**).

A similar process as described above was used for article screening, assessment of study quality, data extraction and data synthesis.

2.3.2 Results

2.3.2.1 Literature Search and identification results

Of the 220 studies identified after removal of duplicate studies, 17 met the inclusion criteria for this review. The included articles were published between the years 2012 to 2017 and contributed to a total of 4,488 women included in our review. Half of the included studies were conducted in the USA (N=9); the others were in Spain (N=2), the United Kingdom and Ireland, Brazil, India, Mexico, Hungary, and Mozambique (N=1 each). Women were usually recruited from obstetric units at a median of 32 weeks (range 23 to 37 weeks). Nine studies (52.9%) recruited only women at preterm (<37 weeks gestation). Some studies mentioned that all included women were admitted into hospital (N=5) while the other studies did not specify. The median maternal age was 31.7 years (range 23 to 34 years) and most women were nulliparous (median 56.3%, range 39.8% to 76%).

2.3.2.2 Quality of studies

The vast majority of studies were of prospective cohort design (N=15), except for two retrospective cohort studies (**Figure 2-3**). All but one of the included studies (N=16) had adequately described population selection, tests and measurements used, and outcomes. Fourteen

of the studies specified masking of the clinicians to the PlGF test results and the technicians to the adverse outcomes. The rate of withdrawal and loss to follow-up were mentioned in only six papers, and only four papers (23.5%) reported sufficient sample size for their study. In total, seven studies were classified as having low risk of bias, and 10 had medium risk of bias.

2.3.2.3 Definition of HDPs

One study recruited women with any confirmed HDPs **(Table 2-5)**.[77] Six studies[57,58,61,78-80] recruited solely women with diagnosed pre-eclampsia, among these, two (N=2 studies) were specifically included women with early-onset pre-eclampsia (GA <34 weeks).[57,79] Ten studies recruited women with suspected pre-eclampsia;[61,73,76,80-86] the rate of confirmed HDPs (chronic and gestational hypertension and the women who went on to have confirmed pre-eclampsia) stated in the studies ranged from 71% to 95%. Some of these studies on suspected pre-eclampsia did not report on the prevalence of confirmed HDPs (N=5 studies).

Chronic hypertension was generally defined as hypertension occurring before pregnancy or before 20 weeks of gestation and gestational hypertension was defined as new-onset hypertension occurring from 20 weeks of gestation. Pre-eclampsia was defined in the studies using international guidelines: ACOG (N=13), ISSHP (N=3), or NHBPEP (N=1) guidelines.

2.3.2.4 Classification of PlGF

PlGF was investigated alone in nine studies and in combination with other angiogenic factors in eight. The cut-off for PlGF recommended as the best threshold for the prediction of adverse outcomes varied in the studies, from ≤0.4 to ≤122 pg/ml; one study used <5th centile for gestational age at testing. Two of these studies specified the conversion of PlGF measures into multiples of median (MOM).[81,82]

In some of the studies, sFlt-1 (pg/ml) was combined with PlGF (pg/ml) as a ratio i.e. sFlt-1/PlGF ratio. Six of these studies used a cut-off of ≥85; other cut-offs ranged from 178 to ≥871. Two of these studies also combined the sFlt-1/PlGF ratio with other clinical variables such as gestational age, proteinuria and systolic blood pressure (sBP) in multivariable models.[84] The other angiogenic factor that was combined with PlGF was sENG (pg/ml), which was combined as a ratio (PlGF/sENG) in one study with a cut-off of ≤0.05 to ≤0.07.

The most commonly used PlGF assay was manufactured by Roche diagnostics (N=7 studies); other studies used the Alere Triage (N=4), R&D systems (N=4), and DRG or the KRYPTOR test platforms (N= 1 study each).

2.3.3 Prediction of maternal outcomes

Four studies[57-59,61] evaluated the use of PlGF for the prediction of adverse maternal outcomes in women with suspected or confirmed pre-eclampsia, mostly based on signs and symptoms of pre-eclampsia (**Table 2-6**). Three studies reported on prediction of composite maternal outcomes, using the sFlt-1/PlGF ratio; the cut-off was 85 in two studies[59,61] and 871in another study.[57] The other study used only PlGF to evaluate the prediction of postpartum haemorrhage (PPH). There were no studies of PlGF alone to predict a composite adverse maternal outcome.

Adverse maternal outcome rates were a median of 8.8% (range 8.2%-9.5%), with median (range) sensitivities of 67.5% (52.1-100) and specificities of 73.7% (51.7-77.9). The only study with both sensitivity and specificity above 70% was by Ghosh *et al*7 for the prediction of PPH. Overall, the LRs were poor with the positive LRs ranging from 2-2.4 and negative LRs from 0.35-0.61.

2.3.4 Prediction of adverse perinatal outcomes

Three studies[58,77,79] reported on the prediction of small-for-gestational age (SGA) infants, stillbirth or neonatal death, and composite neonatal outcomes **(Table 2-6)**. One of the studies evaluated the use of PlGF alone;[77] two studies evaluated the sFlt-1/PlGF ratio and one of the studies added gestational age to the ratio.[58,79]

The median rate of outcomes was 27.5% (range 11.0%-44.3%). The sensitivities in these studies ranged from 36.9-92.8% and specificities from 54.1-84.6%. The AUROCs was reported in one study for the prediction of composite neonatal outcomes in early onset pre-eclampsia, to be 0.75 (0.62–0.88) using sFlt-1/PlGF ratio and 0.89 using sFlt-1/PlGF ratio in combination with GA. Only this study reported a moderate LR- for ruling out adverse neonatal outcomes (LR- 0.13 (95% CI: 0.02–0.91) using sFlt-1/PlGF ratio at a cut-off of >655.

2.3.5 Prediction of timed delivery

Nine studies[71;76-78;80-83;86] reported on the prediction of outcomes related to timing of delivery among women who were preterm, either of birth before 37 weeks' (from a median [range] of 32 weeks [30.6-35]), or the need to deliver the woman (for maternal or fetal reasons) within 7 or 14 days of PlGF testing (from a median [range] of 31 weeks [30.6-32) **(Table 2-6C)**. Seven of 11 studies evaluated the use of PlGF alone; five studies evaluated the sFlt-1/PlGF ratio and two studies evaluated PlGF/ sENG.

The median rate of outcomes was 48% (range 18.0%-68.8%). The sensitivities in these studies ranged from 28-96% and specificities from 55-97.8%. AUROCs were reported in four of these studies and ranged from 0.83 to 0.95. The study with the highest AUROC was by De Oliveira *et.al*[78] for prediction of delivery due to severe PE using sFlt-1/PlGF at cut-off of 85

(AUROC 0.95 (95% CI: 0.92–0.99)). Overall, the studies appeared to have good clinical utility with LR+ ranging from 2.02 to 33.50 and LR- from 0.07 to 0.80.

The study by Chaiworapongsa *et al*[82] showed improvement in the prediction of delivery within 2 weeks for women first presenting at GA<34 weeks, after the combination of PlGF with a ratio either as PlGF/sFlt-1or PlGF/ sENG, compared with using PlGF alone (LR+ from 9.0 (2.3–35) to 22.2 (3.23 – 152.69) and LR- from 0.30 (0.1–0.6) to 0.12 (0.03-0.42)). The best LR– was observed in the study by Chappell *et al*[83] for the prediction of preterm delivery within 14 days for women with suspected pre-eclampsia first presenting at GA<35 weeks, using PlGF only, with a cut-off at <5th centile (0.07 (0.02–0.22)).

2.3.6 Prediction of combined maternal and fetal outcomes

Five studies[61;80;84-85;87] evaluated the use of PlGF as a predictor of combined maternal and fetal outcomes in women with suspected or confirmed pre-eclampsia **(Table 2-3)**. Four of these studies[61;80;84-85] were on women with suspected pre-eclampsia and used sFlt-1/PlGF cut-off of ≥85; one of the four studies also combined the ratio with sBP and proteinuria,[85] and another also evaluated the addition of a clinical multivariable model with 11 variables: race, chronic hypertension, history of renal disease, gravidity (primigravid vs. multigravid), preeclampsia, history, maternal age, smoking status, obesity (BMI >30 kg/m^2), pre-gestational diabetes, clinical diagnosis of preeclampsia, and gestational age at presentation.[84] One study evaluated the prognostic value using PlGF only for women with pre-eclampsia.[87]

The outcome rates ranged from a median of 41.5% (range 28.3-68.8%). The composite outcomes included outcomes such as acute renal failure, thrombocytopenia and pulmonary oedema for maternal outcomes and SGA, stillbirth and neonatal death for fetal outcomes. The median AUROC was 0.81 (range 0.76-0.93). The highest reported sensitivity and specificity was

72.9% (95% CI 59.5-83.3) and 94.0% (95% CI 87.6-97.4), respectively. This study also reported the highest AUROC of 0.93 (0.89–0.97) using sFlt-1/PlGF ratio and had a LR – of 0.29 (0.19-0.44) and a good LR+ of 12.2 (5.8-25.4).

One study investigated adverse outcomes in twin pregnancies[77] and reported an AUROC of 0.75 (0.64–0.86) for the prediction of outcomes in all included women and an AUROC of 0.81 for the prediction of outcomes in women who had enrolled into the study before 34 weeks of gestation.

The study by Salahuddin *et al*[85] which combined sBP and proteinuria reported an AUROC of 0.80 (0.76–0.85); this did not significantly increase when evaluated only in women presenting before GA at 34 weeks (AUROC of 0.89 (95% CI: 0.82–0.95)).

The multivariable model study by Moore *et al*[84] reported a significant increase in AUROC from 0.76 (0.66-0.85) to 0.91 (0.85-0.97) after addition of 11 variables.

All of the studies reported AUROCs ≥0.7 and thus, appeared to have good discriminatory ability.

2.3.7 Discussion

2.3.7.1 Main findings

Prediction of maternal outcomes was generally poor based on LRs, with the best prediction performance observed for of PlGF for PPH.[57] The studies investigating adverse perinatal outcomes were also poor except for only one study with moderate LRs- of 0.13 for ruling out composite adverse fetal outcomes while the studies on timed delivery mostly reported moderate to high LRs+ for ruling in (N= 6/9) and moderate to high LRs- (N= 4/9) for ruling out composite. The discrimination capacity of PlGF in predicting combined adverse maternal and fetal outcomes was high in all studies with AUROC ≥ 0.7; only one of the studies reported LRs, which also had

a high LR of 12 for ruling in disease.[80] However, the extent to which the performances for the combined outcomes were driven by the fetal outcomes is unclear.

2.3.7.2 Comparison with the literature

A review on the accuracy of PlGF along with other angiogenic factors for the prediction of pre-eclampsia reported that although the concentrations of PlGF, sFlt-1 and sENG were significantly altered in pregnancies complicated by pre-eclampsia, these markers in their included studies (N=34) did not show strong prediction of pre-eclampsia independently.[74] This review suggested that the addition of PlGF to multivariable models might be useful in increasing performance. Three of the studies included in our review added PlGF or sFlt-1/PlGF ratio to other variables.[79;84-85] However, there were no significant differences observed in two of these studies upon addition of other factors; one of which added gestational age for the prediction of a composite neonatal outcome,[79] and the other included both sBP and proteinuria[85] for the prediction of a combined maternal and fetal outcome. The third study[84] reported a significant improvement in the prediction of a combined maternal and fetal outcomes upon the inclusion of sFlt-1/PlGF to a clinical multivariable model (from to 0.76 to 0.91). However, this model included 11 other variables with a limited sample size of 78 outcomes. Therefore, the model may have been over-fitted, as the recommended rule of thumb for variable is to have at least ten outcomes per predictive variable assessed. However, it may be worthwhile investigating if the inclusion of other factors to PlGF might improve the prognostic capacity for the prediction of maternal outcomes.

2.3.7.3 Strengths and Limitations

A systematic search strategy was used to identify relevant articles on the use of PlGF as a prognostic factor for women with suspected or confirmed pre-eclampsia in this review, without

any restrictions on language or year of publication. Most studies appropriately reported on quality of study attritions as required by the QUIPS tool, except for inadequate sample size reported and handling of missing data. Therefore, all the included studies were considered to be of good quality in general.

One limitation of this review was that studies recruiting women with suspected pre-eclampsia, in which some of the women did not have any confirmed HDP, were included; although, the reported incidence of any HDP in the included studies ranged from 71%-95%. It was impossible to tease out the prediction ability for only HDPs in the studies recruiting both women with and without HDPs to know if the reported prognostic accuracy would have significantly differed in the women with only HDPs. However, focusing on studies only including women with confirmed HDPs, still had moderate to high LRs reported for timing of delivery and neonatal outcomes but not for adverse maternal outcomes.

Another limitation in this review is that we were unable to assess if PlGF performed better in women at higher risk of adverse outcomes due to limited information on hospital admission, as this information was not provided in majority of the studies. Generally, women who are admitted are considered to be sicker.

Also, there was limited ability to comment on prediction of adverse maternal outcomes given that there were few informative studies. Therefore, it is difficult to make strong inferences about the use of PlGF to determine prognosis for maternal outcomes in women with HDP.

2.4 Conclusion

The fullPIERS model for the prediction composite maternal outcomes in women with pre-eclampsia and superimposed pre-eclampsia was the best performing multivariable model in our review. In our systematic review of PlGF as a prognostic biomarker, PlGF showed promising

results for the prediction of delivery and also for the prediction of combined maternal and fetal outcomes in women with HDPs. However, its usefulness for the prediction of solely maternal outcomes was not demonstrated in the included studies although there were limited studies. This suggests the needs for more studies to evaluate the ability of PlGF to predict adverse maternal outcomes from HDPs.

Table 2-1 Inclusion and exclusion criteria for predictors of adverse maternal outcomes

Selection criteria	Inclusion	Exclusion
Population	Studies recruiting patients with any HDP – gestational hypertension, chronic hypertension, pre-eclampsia, super-imposed pre-eclampsia and HELLP syndrome	Studies not recruiting women with HDP
Intervention/ Study design	Studies reporting risk prediction or prognosis tests for adverse maternal outcomes resulting from HDPs or studies that provide data that can be used to calculate these tests. Prognostic tests include at least one of the following: area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, likelihood ratios, negative or positive predictive value.	Studies that do not report prognosis tests OR studies not presenting sufficient data for calculation
Comparators	None	
Outcomes	Primary outcome measures were studies including any PIERS adverse maternal outcomes in **Appendix B** and online (https://pre-empt.cfri.ca/monitoring/fullpiers) or/and Postpartum haemorrhage (PPH); disseminated intravascular coagulation (DIC)	Studies without outcome measures OR Studies that report on combined fetal and maternal outcome and the prediction of maternal outcome cannot be separated.
Language	None	
Publication year limit	None	

Table 2-2 Study characteristics for review on maternal outcomes

Author, Year	Study design/ Quality score	Country(ies)	Time period	Type of HDP	Maternal characteristics	Outcome(s)	Prediction method	Tests
Ankumah et al. 2014	Multicentre RCT 7	USA	May 1991-Jun 1995	Mild chronic hypertension	Mean age: 30 Mean GA: 19.8 Nulliparous: 18.2%	Placental abruption	Univariable	Blood pressure
Aziz et al. 2011	Retrospective review 4	India	Jan 2005-Dec 2009	HELLP syndrome	Mean age: 26.6 Mean GA: 32.9 Nulliparity: 62.2%	Composite: DIC, ARF, PPH, Placental abruption, cerebral or pulmonary edema, liver infarcts or rupture, or a subcapsular liver hematoma or maternal death	Univariable	Epigastric pain; vomiting; headache; visual symptoms; non-specific viral symptoms; platelets; AST; ALT; LDH
Ben Salem et al. 2003	Case-control 4	Tunisia	Jan 1995-Jun 2000	Pre-eclampsia	Mean age: 30 Nulliparity: 38.3%	Eclampsia	Univariable	sBP; dBP; headache; visual symptoms; hyperreflexia; proteinuria; uric acid; serum creatinine; AST
Bouzari et al. 2014	Retrospective cohort 6	Iran	2000-2010	Pre-eclampsia	Mean age: 29.5	Placental abruption	Univariable	24h proteinuria
Chan et al. 2005	Retrospective cohort 5	Australia	1998-2001	Pre-eclampsia	Mean Age: 30 Nulliparity: 73%	Composite: renal insufficiency (creatinine >90 µmol/L), liver disease (AST >40 U/L), cerebral irritation (hyperreflexia with clonus or repeated visual scotomata, requiring magnesium sulphate) and thrombocytopenia (platelets <150* 10^9/L)	Multivariable logistic regression	Spot urine PRCR and maternal age at diagnosis

Author, Year	Study design/ Quality score	Country(ies)	Time period	Type of HDP	Maternal characteristics	Outcome(s)	Prediction method	Tests
Gangaram et al. 2009	Prospective cohort 6	South Africa	January 2006	Gestational hypertension and pre-eclampsia	Median GA: 33 Nulliparity: 12.9%	Composite: Placental abruption, eclampsia	Univariable	Spot urine ACR
Ghosh et al. 2012	Prospective cohort 5	India	Mar 2009-Jun 2011	Early onset preeclampsia <34 weeks	GA presentation: 23 Median Age: 23 Nulliparity: 76%	Postpartum hemorrhage (PPH) defined as a blood loss of >1500 ml and/or the need for a blood transfusion, for a caesarian delivery and >1000 ml and/or the need for a blood transfusion, in case of a vaginal birth	Univariable	Serum PlGF
Girling et al. 1997	Prospective cross-sectional 5	UK	-	Pre-eclampsia	Mean age: 29	Composite: acute renal failure requiring dialysis, profound oliguria needing central venous pressure monitoring and renal support, and spontaneous pulmonary oedema	Multivariable	Abnormal liver function tests (AST, ALT, bilirubin, GGT)
Hall et al. 2002	Prospective cohort 5	South Africa	Apr 1992 – Mar 1997	Early onset, severe pre-eclampsia	Mean Age: 27 Mean GA: 29.5	Placental abruption, ascites, pulmonary edema, eclampsia	Univariable	24h proteinuria
Kozic et al. 2011	Multicentre prospective review 6	Canada, Australia, New Zealand, UK	Sep 2003 – Jan 2010	Pre-eclampsia	Mean: 31 Nulliparity: 71.1%	PIERS Composite	Univariable	AST, ALT, LDH, albumin, total bilirubin, INR
Laskin et al. 2011	Multicentre prospective cohort 5	Canada, New Zealand, Australia, UK	Sep 2003 – Jan 2010	Pre-eclampsia	Median age: 31 Median GA: 33.7 Nulliparity: 72.6%	PIERS composite outcome	Univariable	Platelet count

Author, Year	Study design/ Quality score	Country(ies)	Time period	Type of HDP	Maternal characteristics	Outcome(s)	Prediction method	Tests
Leaños-Miranda et al. 2013	Prospective cohort 5	Mexico	-	Pre-eclampsia	GA presentation: 32 Mean Age: 28.3 Nulliparous: 43.5%	Composite: maternal mortality and any of the following serious maternal morbidities: hepatic hematoma or rupture (confirmed by ultrasound or laparotomy), pulmonary edema (clinical diagnosis and with radiographic confirmation), need for positive inotropic support, intubation (other than solely for caesarean section), acute renal failure (creatinine ≥198 μmol/L), and placental abruption (clinical or pathological).	Univariable	Serum sFlt-1/PlGF ratio
Livingston et al. 2014	Multicentre prospective cohort 7	Canada, New Zealand, Australia, UK	Sept 2003 – Dec 2011	Pre-eclampsia	Median Age: 31 Median GA: 35 Nulliparity: 73.8%	PIERS composite outcome	Univariable	Uric acid
Millman et al. 2011	Multicentre prospective cohort 5	Canada, New Zealand, Australia, UK	Sept 2003 – Jan 2010	Pre-eclampsia	Median Age: 31 Median GA:34.1 Nulliparity: 89.2%	PIERS composite outcome	Univariable and multivariable	SpO2; chest pain and/or dyspnea
Palomaki et al. 2015	Prospective cohort 8	USA	Jul 2009 – Jun 2012	HDP	Mean GA: 30	Composite (Placental abruption, Acute renal failure, DIC, Pulmonary edema)	Univariable	sFlt-1/PlGF ratio

Author, Year	Study design/ Quality score	Country(ies)	Time period	Type of HDP	Maternal characteristics	Outcome(s)	Prediction method	Tests
Payne et al. 2011	Multicentre prospective cohort 7	Canada, New Zealand, Australia, UK	Sept 2003 – Jan 2010	Pre-eclampsia	Median age: 31 Median GA: 36 Nulliparity: 71.4%	PIERS composite outcome	Univariable	Proteinuria (dipstick, spot PRCR, 24h protein)
Payne et al. 2014	Multicentre prospective cohort 10	Uganda, South Africa, Brazil, Pakistan	Jul 2008 – Mar 2012	Any HDP	Median age: 28 Median GA: 35.3 Nulliparity: 46.1%	PIERS composite outcome	Multivariable logistic regression	Parity, GA at assessment, chest pain/dyspnea, headache/visual disturbances, vaginal bleeding with abdominal pain, sBP, dipstick proteinuria
Payne et al. 2015	Multicentre prospective cohort 8	South Africa, Pakistan	Jan 2011 – Dec 2013	All HDP	Mean age: 28 Nulliparity: 47.7%	PIERS composite outcome	Multivariable logistic regression	Parity, GA at assessment, Chest pain/dyspnea, Headache/visual disturbances, Vaginal bleeding with abdominal pain, sBP, dipstick proteinuria and SpO_2
Rana et al. 2013	Prospective cohort 3	USA	Jul 2009 – Oct 2010	Pre-eclampsia	Mean Age: 32 Mean GA: 35 Nulliparity: 70.1%	Composite: abnormal liver function test and platelets, placental abruption, pulmonary edema, cerebral hemorrhage, seizure (in the absence of an underlying seizure disorder), acute renal failure (creatinine41.5mg/dL) or maternal death	Spearman rank correlation	sFlt1/PlGF ratio
Romero et al. 1988	Retrospective review	USA	Jan 1981 – Sep 1984	GH, Pre-eclampsia and	Mean age: 25.5 Nulliparity: 71.6%	Pulmonary edema, eclampsia	Univariable	AST

Author, Year	Study design/ Quality score	Country(ies)	Time period	Type of HDP	Maternal characteristics	Outcome(s)	Prediction method	Tests
	6			superimposed pre-eclampsia				
Saleh et al. 2016	Prospective 6	Netherlands	Sep 2011 – Aug 2013	Pre-eclampsia	Mean age: 32 Mean GA: 30	Composite: Pulmonary edema, Acute renal failure	Binary logistic analysis	sFlt1/PlGF ratio
Scazzochio et al. 2013	Prospective 7	Spain	Sep 2010 – Sep 2012	Severe preeclampsia <34 weeks	Mean age: 32 Mean GA: 30.3 Primiparity: 58.2%	Composite: Acute renal failure and pulmonary edema	Univariable	Maternal Neutrophil Gelatinase-Associated Lipocalin (NGAL)
Schiff et al. 1996	Retrospective review 5	USA	Jan 1990 – Dec 1994	Severe pre-eclampsia	Mean Age: 23 Mean GA: 31.2 Nulliparity: 50%	Placental abruption	Univariable	24h urine proteinuria
von Dadelszen et al. 2011	Multicentre prospective cohort 10	Canada, New Zealand, Australia, UK	Sep 2003 – Jan 2010	Pre-eclampsia	Median age: 31 Median GA: 35 Nulliparity: 71.3%	PIERS composite outcome	Multivariable logistic regression	GA, Chest pain or dyspnoea, oxygen saturation, platelet count, creatinine, and AST
Witlin et al. 1999	Prospective 5	USA	Mar 1992 – Jan 1997	Severe pre-eclampsia	Mean Age: 22.8	Eclampsia, placental abruption	Univariable	Mean arterial pressure (MAP), Platelet count
Yassaee et al. 2003	Cohort 3	Iran	1986-2001	Severe pre-eclampsia	-	Maternal mortality, eclampsia	Univariable	Uric acid
Yen et al. 2011	Multicentre prospective cohort 5	Canada, New Zealand, Australia, UK	Sep 2003 – Jul 2009	Pre-eclampsia	Mean Age: 32 Mean GA: 34.4 Nulliparity: 71.3%	PIERS composite outcome	Univariable	Nausea and vomiting; Headache; Visual disturbances; Right upper quadrant or epigastric pain; Abdominal pain or vaginal bleeding; Chest pain or dyspnea
Yucesoy et al. 2005	Prospective cohort 5	Turkey	-	HELLP syndrome	Mean age: 26.5 Mean GA: 31.5 Nulliparity: 29.5%	Eclampsia, placenta abruption, DIC, acute renal failure, maternal mortality	Univariable	Platelet count

Table 2-3 Tests accuracy for prediction of adverse maternal outcomes in HDP, N = 28 studies

Author, Year	Test / cut-off	Outcome	Total (outcome rate %)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUROC (95% CI)
UNIVARIABLE TESTS								
Signs and/or symptoms								
Aziz et al. 2011	Headache	Composite	74 (27%)	30.0 (11.9-54.3)	46.3 (32.6-60.4)	0.6 (0.3-1.1)	1.5 (1.0-2.3)	0.40 (0.20-0.50)
Ben Salem et al. 2003	Headache	Eclampsia	120 (34.2%)	97.6 (85.6-99.9)	26.6 (17.6-37.9)	1.3 (1.2-1.5)	0.1 (0-0.7)	-
Yen et al. 2011	Headache	PIERS Composite	2020 (7.1%)	-	-	-	-	0.535 (0.47-0.58)
Aziz et al. 2011	Vomiting	Composite	74 (27%)	10.0 (1.2-31.7)	77.9 (64.4-88.0)	0.5 (0.1-1.8)	1.2 (0.9-1.4)	0.40 (0.30-0.50)
Yen et al. 2011	Nausea/vomiting	PIERS Composite	2020 (7.1%)	-	-	-	-	0.54 (0.48-0.60)
Ben Salem et al. 2003	Visual symptoms	Eclampsia	120 (34.2%)	85.4 (70.1-93.9)	65.8 (54.2-75.9)	2.5 (1.8-3.5)	0.2 (0.1-0.5)	-
Yen et al. 2011	Visual symptoms	PIERS Composite	2020 (7.1%)	-	-	-	-	0.50 (0.45-0.56)
Yen et al. 2011	Abdominal pain or vaginal bleeding	PIERS Composite	2020 (7.1%)	-	-	-	-	0.57 (0.47-0.67)
Aziz et al. 2011	Epigastric pain	Composite	74 (27%)	10.0 (1.2-31.7)	70.4 (56.4-82.0)	0.3 (0.1-1.3)	1.3 (1.0-1.6)	0.4 (0.3-0.5)
Yen et al. 2011	RUQ or epigastric pain	PIERS Composite	2020 (7.1%)	-	-	-	-	0.605 (0.545-0.664)
Millman et al. 2011	Chest pain and/or dyspnoea	PIERS Composite	1534 (6.1%)	-	-	-	-	0.59 (0.52-0.65)
Millman et al. 2011	Chest pain and/or dyspnoea	Non-respiratory PIERS Composite	1534 (4.4%)	-	-	-	-	0.53 (0.45-0.60)
Yen et al. 2011	Chest pain or dyspnoea	PIERS Composite	2020 (7.1%)	-	-	-	-	0.58 (0.52-0.64)
Millman et al. 2011	SpO_2 <93%	PIERS Composite	1534 (6.1%)	-	-	-	-	0.71 (0.65-0.77)

Author, Year	Test / cut-off	Outcome	Total (outcome rate %)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUROC (95% CI)
Millman et al. 2011	SpO_2 <93%	Non-respiratory PIERS Composite	1534 (4.4%)					0.64 (0.57-0.71)
Aziz et al. 2011	Non-specific viral symptoms	Composite	74 (27%)	65.0 (40.8-84.6)	87.0 (75.1-94.6)	5.0 (2.3-10.7)	0.4 (0.2-0.7)	0.80 (0.60-0.90)
Ben Salem et al. 2003	Vivid deep tendon reflexes	Eclampsia	120 (34.2%)	97.6 (85.6-99.9)	46.8 (35.6-58.3)	1.8 (1.5-2.3)	0.1 (0.0-0.4)	-
Blood pressure (BP)								
Ben Salem et al. 2003	sBP ≥ 160 mmHg	Eclampsia	120 (34.2%)	92.7 (79.0-98.1)	24.1 (15.4-35.2)	1.2 (1.0-1.4)	0.3 (0.1-1.0)	-
Ankumah et al. 2014	sBP and/or dBP >140/90 mmHg	Placental abruption	759 (1.4%)	36.4 (12.4-68.4)	62.8 (59.2-66.3)	1.0 (0.4–2.1)	1.0 (0.6–1.9)	-
Witlin et al. 1999	MAP >105 mmHg	Eclampsia	445 (9.0%)	92.5 (78.5-98.0)	3.2 (1.8-5.6)	1.0 (0.9–1.0)	2.3 (0.9–8.0)	-
Witlin et al. 1999	MAP >105 mmHg	Placental abruption	445 (7.2%)	87.5 (70.1-95.9)	2.2 (1.1-4.2)	0.9 (0.8–1.0)	5.7 (1.9–17.8)	-
Proteinuria								
Ben Salem et al. 2003	Dipstick >3+	Eclampsia	120 (34.2%)	85.3 (70.1-93.9)	53.2 (41.7-64.4)	1.8 (1.4-2.4)	0.3 (0.1-0.6)	-
Ben Salem et al. 2003	24h urine >3g/d	Eclampsia	120 (34.2%)	36.6 (22.6-53.1)	91.1 (82.0-96.1)	4.1 (1.8-9.3)	0.7 (0.6-0.9)	-
Bouzari et al. 2014	24h urine >1.75g/d	Placental abruption	289 (5.9%)	94.1 (69.2-99.7)	63.7 (57.5-69.3)	2.6 (2.1–3.1)	0.1 (0.0–0.6)	0.777
Gangaram et al. 2009†	Spot urine ACR ≥300mg/g	Composite	155 (2.6%)	0	55.0 (46.7-63.0)	-	1.8 (1.8-1.8)	
Hall et al. 2002	24h urine increased by ≥2g	Placental abruption	74 (13.5%)	30.0 (8.1-64.6)	59.4 (46.4-71.2)	0.7 (0.3-2.0)	1.2 (0.8-1.8)	-
Hall et al. 2002	24h urine increased by ≥2g	Ascites	74 (10.8%)	62.5 (25.9-89.8)	63.6 50.8-74.9)	1.7 (0.9-3.2)	0.6 (0.2-1.5)	-
Hall et al. 2002	24h urine increased by ≥2g	Pulmonary edema†	74 (1.4%)	0 (0-94.5%)	60.3 (48.1-71.3)	-	1.7 (1.6-1.7)	
Hall et al. 2002	24h urine increased by ≥2g	Eclampsia†	74 (1.4%)	100 (5.5-100)	61.6 (49.5-72.6)	2.6 (1.9-3.5)	-	-
Payne et al. 2011	Dipstick	PIERS Composite	2002 (5.3%)	-	-	-	-	0.55 (0.49-0.61)

Author, Year	Test / cut-off	Outcome	Total (outcome rate %)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUROC (95% CI)
Payne et al. 2011	Spot urine PRCR	PIERS Composite	2002 (5.3%)	-	-	-	-	0.48 (0.42-0.55)
Payne et al. 2011	24hr urine	PIERS Composite	2002 (5.3%)	-	-	-	-	0.55 (0.47-0.63)
Schiff et al. 1996	24h urine increased by ≥2g	Placental abruption	2002 (5.3%)	40 (7.3-83.0)	63.9 (50.6-75.5)	1.1 (0.4-3.4)	0.9 (0.5-2.0)	-
Laboratory tests								
Aziz et al. 2011 2011	Platelets ≤100 × 10^5/L	Composite	74 (27%)	70.0 (45.7-88.1)	20.4 (10.6-33.5)	0.9 (0.6-1.2)	1.5 (0.6-3.4)	0.40 (0.30-0.60)
Laskin et al. 2011	Platelets ≤100 × 10^9/L	PIERS Composite	1405 (10.8%)	15.8 (10.6-22.8)	92.2 (90.5-93.6)	2.0 (1.3–3.1)	0.9 (0.9–1.0)	-
Witlin et al. 1999	Platelets <60,000/mm^3	Placental abruption	445 (7.2%)	37.5 (21.7-56.3)	85.0 (81.1-88.2)	2.5 (1.5–4.1)	0.7 (0.6–1.0)	-
Yucesoy et al. 2005	Platelets <50,000/mm^3	Eclampsia	44 (29.5%)	38.5 (15.1-67.7)	64.5 (45.4-80.2)	1.1 (0.5–2.5)	1.0 (0.6–1.5)	-
Yucesoy et al. 2005	Platelets <50,000/mm^3	Placenta abruption	44 (11.4%)	40.0 (7.3-83.0)	64.1 (47.1-78.3)	1.1 (0.4–3.5)	0.9 (0.4–2.0)	-
Yucesoy et al. 2005	Platelets <50,000/mm^3	Disseminated intravascular coagulation	44 (18.2%)	75.0 (35.6-95.5)	72.2 (54.6-85.2)	2.7 (1.4–5.2)	0.3 (0.1–1.2)	-
Yucesoy et al. 2005	Platelets <50,000/mm^3	Acute renal failure	44 (15.9%)	71.4 (30.3-94.9)	70.3 (52.8-83.6)	2.4 (1.2–4.8)	0.4 (0.1–1.3)	-
Yucesoy et al. 2005	Platelets <50,000/mm^3	Maternal mortality	44 (9.1%)	25.0 (1.3-78.1)	62.5 (1.3-76.8)	0.7 (0.1–3.8)	1.2 (0.6–2.2)	-
Kozic et al. 2011	INR	PIERS Composite	2008 (5.1%)	-	-	-	-	0.65 (0.58-0.71)
Ben Salem et al. 2003	Creatinine >100µmol/L	Eclampsia	120 (34.2%)	39.0 (24.6-55.5)	81.0 (70.3-88.6)	2.1 (1.1-3.7)	0.8 (0.6-1.0)	-
Ben Salem et al. 2003	Uric acid ≥350 µmol/L	Eclampsia	120 (34.2%)	82.9 (67.4-92.3)	65.8 (54.2-92.3)	2.4 (1.7-3.4)	0.3 (0.1-0.5)	-
Livingston et al. 2014	Uric acid >345 µmol/L	PIERS Composite	1487 (13.3%)	80.2 (70.8- 87.6)	28.2 (25.9-30.7)	1.1 (1.0–1.2)	0.7 (0.5–1.0)	0.62 (0.56-0.69)
Yassaee et al. 2003†	Uric acid ≥6mg/dL	Maternal mortality†	103 (8.7%)	100 (62.9-100)	53.2 (42.6-63.4)	2.1 (1.7–2.7)	0	-
Yassaee et al. 2003†	Uric acid ≥6mg/dL	Eclampsia	103 (12.6%)	92.3	54.4	2.0	0.1	-

Author, Year	Test / cut-off	Outcome	Total (outcome rate %)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUROC (95% CI)
				(62.1-99.6)	(43.6-64.9)	(1.5–2.7)	(0–0.9)	
Aziz et al. 2011	ALT ≥70 IU/L	Composite	74 (27%)	55.0 (31.5-76.9)	25.9 (15.0-39.7)	0.7 (0.5-1.1)	1.7 (0.9-3.4)	0.4 (0.3-0.5)
Kozic et al. 2011	ALT	PIERS Composite	2008 (5.1%)	-	-	-	-	0 .73 (0 .67-0.79)
Aziz et al. 2011	AST ≥70 IU/L	Composite	74 (27%)	60.0 (36.1-80.9)	42.6 (29.2-56.8)	1.1 (0.7-1.6)	0.9 (0.5-1.8)	0.50 (0.40-0.60)
Ben Salem et al. 2003	AST >30 IU/L	Eclampsia	120 (34.2%)	63.4 (46.9-77.4)	70.9 (59.4-80.3)	2.2 (1.4-3.3)	0.5 (0.4-0.8)	-
Kozic et al. 2011	AST	PIERS Composite	2008 (5.1%)	-	-	-	-	0.73 (0.67-0.79)
Romero et al. 1988	AST 2SD above mean	Pulmonary edema	275 (1.1%)	66.7 (12.5-98.2)	79.4 (74.0-84.0)	3.2 (1.4–7.5)	0.4 (0.1–2.1)	-
Romero et al. 1988	AST 2SD above mean	Eclampsia	275 (2.5%)	71.4 (30.3-94.9)	80.2 (74.8-84.7)	3.6 (2.1–6.1)	0.4 (0.1–1.2)	-
Aziz et al. 2011	LDH ≥600 IU/L	Composite	74 (27%)	75.0 (50.9-91.3)	55.6 (41.4-61.9)	1.7 (1.1-2.5)	0.5 (0.2-1.0)	0.7 (0.6-0.8)
Kozic et al. 2011	LDH	PIERS Composite	2008 (5.1%)	-	-	-	-	0.74 (0.68-0.81)
Kozic et al. 2011	Serum albumin	PIERS Composite	2008 (5.1%)	-	-	-	-	0.63 (0.57-0.69)
Kozic et al. 2011	Total bilirubin	PIERS Composite	2008 (5.1%)	-	-	-	-	0.68 (0 61-0.74)
Biomarkers								
Ghosh et al. 2012	Serum PlGF <122pg/mL	Postpartum hemorrhage (PPH)	766 (8.7%)	73.1 (60.7-82.9)	76.7 (73.3-79.7)	3.14 (2.57-3.82)	0.35 (0.24-0.52)	-
Leaños-Miranda et al. 2013	Serum sFlt-1/PlGF ratio ≥871	Composite	501 (9.6%)	52.1 (37.4-66.5)	77.9 (73.8-81.6)	2.36 (1.71-3.26)	0.61 (0.46-0.83)	-
Palomaki et al. 2015	Serum Flt-1/PlGF ratio >85	Composite	237 (8.9%)	61.9 (38.7–81.0)	69.4 (62.8–75.4)	2.0 (1.4–3.0)	0.5 (0.3–1.0)	-
Rana et al. 2013†	Serum Flt-1/PlGF ratio≥85	Composite	97 (8.2%)	100 (59.7–100)	51.7 (40.9–62.3)	2.1 (1.7–2.6)	∞	-

Author, Year	Test / cut-off	Outcome	Total (outcome rate %)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUROC (95% CI)
Saleh et al. 2016†	Serum Flt-1/PlGF ratio≥85	Composite	62 (9.7%)	100 (51.7–100)	10.7 (4.4–22.6)	1.1 (1.0–1.2)	-	-
Scazzochio et al. 2013	Maternal NGAL >100ng/mL	Composite	67 (17.9%)	41.7 (16.5-71.4)	65.5 (51.3-77.4)	1.2 (0.6-2.6)	0.9 (0.5-1.5)	-
MULTIVARIABLE TESTS, N=6 studies								
Chan et al.† 2005	Spot urine PRCR >500 and maternal age >35 years	Composite	321 (34%)	10.2 (5.4–17.9)	100 (97.8–100)	-	0.9 (0.8–1.0)	0.67 (0.55-0.71)
Girling et al. 1997†	AST 30 nmol/L ALT 32 nmol/L Bilirubin 14nmol/L	Composite	35 (20%)	100 (56.1-100)	57.1 (37.4-75.0)	2.3 (1.5–3.6)	-	-
Millman et al. 2011	Chest pain and/or dyspnoea and SpO_2	PIERS Composite	1534 (6.1%)	-	-	-	-	0.73 (0.67-0.78)
Payne et al. 2014	miniPIERS model ǂ 25% predicted probability	PIERS Composite	2081 (12.5%)	41.4 (35.4–47.6)	91.9 (90.5–93.1)	5.1 (4.1–6.3)	0.6 (0.6–0.7)	0.79 (0.74–0.80)
Payne et al. 2015	miniPIERS model ǂ and SpO_2, 25% predicted probability	PIERS Composite	852 (17.3%)	49.6 (40.3–58.8)	91.5 (89.2–93.4)	5.9 (4.3–7.9)	0.6 (0.5–0.7)	0.81 (0.76-0.86)
von Dadelszen et al. 2011	GA, chest pain or dyspnea, SpO_2, platelet count, creatinine and AST; 30% predicted probability	PIERS Composite	2023 (5%)	44.9 (34.5–55.3)	98.4 (97.6-98.9)	26.5 (17.4–40.2)	0.6 (0.5–0.7)	0.88 (0.84–0.92)

Table 2-4 Inclusion and exclusion criteria for Systematic Review on PlGF

Selection criteria	Inclusion	Exclusion
Population	Studies recruiting women with any suspected or confirmed HDPs	Studies not recruiting women with HDPs
Intervention/ Study design	Studies reporting risk prediction or prognosis tests for adverse maternal or fetal outcomes resulting from HDPs or studies that provide data that can be used to calculate these tests. Prognostic tests include at least one of the following: area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, or likelihood ratios.	Studies that do not report prognosis tests or studies not presenting sufficient data for calculation
Comparators	None	
Outcomes	Adverse maternal outcome measures included the PIERS adverse outcomes (https://pre-empt.cfri.ca/monitoring/fullpiers) or/and Postpartum haemorrhage; disseminated intravascular coagulation (DIC) Adverse fetal outcomes included any of the following: Stillbirth, neonatal death, or preterm delivery <37+^{0} week or iatrogenic delivery within one or two weeks, or significant neonatal morbidity (abnormal umbilical artery Doppler (absent or reverse flow), 5-min Apgar score<7, umbilical artery pH at birth<7.00, intrauterine growth restriction (IUGR) or small-for-gestational age (SGA) birth weight (≤10th percentile for gestational age), NICU admission, intraventricular hemorrhage ≥Grade 3, any grade of periventricular leukomalacia, hypoxic ischemic encephalopathy, necrotizing enterocolitis, acute renal failure or cardiac failure requiring inotropic agents).	Studies without outcome measures.
Language	None	
Publication year limit	None	

Table 2-5 Characteristics of included studies in PlGF review, N = 17 studies

Author, Year	Study design Country	Type of HDP, Time period	Maternal characteristics	Outcome(s)	Inclusion/ Exclusion criteria	Quality Score
Álvarez-Fernández 2016	Retrospective full-blinded cohort study Spain	Suspected PE Admitted at the Obstetric triage January 2010 to March 2014	GA presentation: 31 Median Age: 34 Nulliparity: 75.9%	Delivery within the first week since clinical presentation for **GA<34 weeks**	**Inclusion:** one or more of: high BP, proteinuria, abnormal uterine artery Doppler, symptoms of PE. **Exclusion:** GA < 20 weeks, antiphospholipid syndrome, systemic lupus erythematosus, confirmed PE diagnosis before presentation at triage	5
Chaiworapongsa 2011	Retrospective cohort USA	Suspected PE Presented to the obstetrical triage unit (year not stated)	GA presentation: 33 Median Age: 26 Nulliparity: 42.4%	i) Preterm delivery due to severe PE ii) PE delivered within 2 weeks for GA<34 weeks	**Inclusion:** GA 20–36^{+6} weeks; signs/ symptoms of PE. **Exclusion:** known major fetal or chromosomal anomaly, multiple gestations.	6
Chaiworapongsa 2014	Prospective cohort USA	Suspected PE Presented to the obstetrical triage unit July 2010 to March 2011				8
Chappell 2013	Prospective observational study United Kingdom and Ireland	Suspected PE Presented at maternity units January 2011 to February 2012	GA presentation: 31 Median Age: 31.9 Nulliparity: 43%	Confirmed preeclampsia within 14 days for GA < 35 weeks	**Inclusion:** symptoms or signs of PE, GA 20^{+0} to 40^{+6} weeks, singleton or twin pregnancy, were ≥16 years of age. **Exclusion:** Confirmed PE diagnosis before enrollment	9

Author, Year	Study design Country	Type of HDP, Time period	Maternal characteristics	Outcome(s)	Inclusion/ Exclusion criteria	Quality Score
De Oliveira 2013	Cohort study Brazil	Early-onset PE (GA <36 weeks) Presented at obstetrics units March 2011 to March 2012	GA presentation: 35 Median Age: 26 Nulliparity: 59.1%	Delivery due to PE	**Inclusion:** hospital admission with PE or SIPE (ISSHP definition). **Exclusion: -**	5
Ghosh 2012	Prospective cohort study India	Early onset PE (GA <34 weeks) March 2009 and June 2011	GA presentation: 23 Median Age: 23 Nulliparity: 76%	PPH and/or the need for a blood transfusion, for a caesarian delivery and >1000 ml and/or the need for a blood transfusion, in case of a vaginal birth)	**Inclusion:** PE (ISSHP definition) **Exclusion:** known history of clotting or bleeding disorders	6
Go´Mez-Arriaga 2014	Prospective cohort study Spain	Early-onset PE (GA <34 weeks) Attending maternity centre November 2007 to December 2012	GA diagnosis: 29.3 Mean Age: 31.7 Nulliparity: 49%	Composite Neonatal outcome: 5-min Apgar score<7, umbilical artery pH at birth<7.00, fetal or neonatal death or significant neonatal morbidity (IVH ≥Grade 3, any grade of PVL, HIE, NEC, acute renal failure or cardiac failure requiring inotropic agents).	**Inclusion:** PE (ISSHP definition), GA< 34^{+0} weeks **Exclusion:** Pregnancies with fetal congenital anomalies and multiple pregnancy	6
Leaños-Miranda 2013	Prospective cohort study Mexico	PE Hospital admission -	GA presentation: 32 Mean Age: 28.3 Primigravida: 43.5%	Composite maternal outcomes: maternal mortality, hepatic hematoma or rupture, pulmonary edema, need for PIS, intubation, acute renal failure, placental abruption. Composite fetal/neonatal outcomes: stillbirths, neonatal death and SGA	**Inclusion:** Diagnosed with PE (ACOG definition) **Exclusion: -**	5

Author, Year	Study design Country	Type of HDP, Time period	Maternal characteristics	Outcome(s)	Inclusion/ Exclusion criteria	Quality Score
Molvarec 2013	Observational study Hungary	Any HDP Presenting at Hospital May 2008 to October 2010.	GA presentation: 32 Median Age: 33 Primigravida: 57.3%	SGA and preterm delivery	**Inclusion:** preeclampsia (ACOG and NHBPEP), HELLP syndrome, superimposed preeclampsia, chronic hypertension, and gestational hypertension **Exclusion:** Not stated	7
Moore 2012	Prospective cohort study USA	Suspected PE Presenting at Hospital January 2007 to September 2010	GA presentation: 37 Mean Age: 30 Primigravida: 2.2%	Composite maternal outcomes: AST or ALT >70 U/l, thrombocytopenia, hemolysis, oliguria, acute renal failure, seizure, pulmonary edema, cerebral hemorrhage, or maternal death. Composite neonatal outcomes: early preterm birth (<34 weeks), placental abruption, neonatal, ICU admission, or fetal/neonatal death.	**Inclusion:** Any of: elevated BP, proteinuria, symptoms of PE. **Exclusion:** Women < 18 years of age, unable to understand English, or unable to provide informed consent.	6
Palomaki 2015	Prospective USA	Suspected preterm PE (GA ≤34 weeks) presented to the obstetric triage unit July 2009 to June 2012	Mean GA: 30 - -	Composite maternal outcomes: Placental abruption, Acute renal failure, DIC, Pulmonary edema	**Inclusion:** preeclampsia (ACOG) ≤34^{+0} weeks, singleton pregnancies **Exclusion: -**	6

Author, Year	Study design Country	Type of HDP, Time period	Maternal characteristics	Outcome(s)	Inclusion/ Exclusion criteria	Quality Score
Rana 2012	Prospective cohort study USA	Suspected Pre-eclampsia Presented to the obstetric triage unit July 2009 to October 2010	Median Age: 32 Median GA: 31 Nulliparity: 56.3%	Composite maternal outcomes: hypertension plus one of the following: elevated AST or ALT, thrombocytopoenia, DIC, abruption, pulmonary edema, cerebral hemorrhage, seizure, acute renal failure (creatinine 114.4 Umol/L), or maternal death. Composite fetal/neonatal outcomes: iatrogenic delivery for HDP, SGA, abnormal UAD, fetal death, and neonatal death.	**Inclusion:** Any: elevated BP, proteinuria, or PE symptoms. **Exclusion:** Multiple pregnancy	8
Rana 2012	Prospective cohort study USA Single	Suspected PE with **twins** Presented to the obstetric triage unit July 2009 to August 2011	Median Age: 33 Median GA: 33.9 Nulliparity: 74.7%		**Inclusion:** Any: elevated BP, proteinuria, or PE symptoms, twins **Exclusion:** singleton pregnancy	8
Rana 2013	prospective cohort study USA Single	PE Presented to the obstetric triage unit July 2009 to October 2010	Mean Age: 32 Mean GA: 35 Nulliparity: 70.1%		**Inclusion:** PE (ACOG definition) **Exclusion:** Multiple pregnancy	6
Salahuddin 2016	prospective cohort study USA Single	Suspected PE, **GA<34** July 2009 to October 2010	Median Age: 32 Mean GA: 31.9 Nulliparity: 55.5%	Composite maternal outcomes: hypertension plus one of: elevated AST or ALT, thrombocytopoenia, DIC, abruption, pulmonary edema, cerebral hemorrhage, seizure, acute renal failure, or maternal death. Composite fetal/neonatal outcomes included iatrogenic delivery indicated for HDPs.	**Inclusion:** elevated BP, proteinuria or any PE symptoms **Exclusion:** Multiple pregnancy	6

Author, Year	Study design Country	Type of HDP, Time period	Maternal characteristics	Outcome(s)	Inclusion/ Exclusion criteria	Quality Score
Ukah 2017	Prospective cohort study Mozambique	Suspected PE Presenting at antenatal clinics August 2014 to February 2015	Median Age: 24.3 Mean GA: 33.1 Nulliparity: 39.8%	Preterm delivery	**Inclusion:** ≥16 years of age, GA ≥20^{+0} weeks, and PE symptoms or hypertension. **Exclusion: -**	7
Woelkers 2016	Prospective observational study USA	Suspected PE Presented to the obstetric triage unit May 2010 to March 2012	-	Preterm delivery or delivery within 14 days	**Inclusion:** symptoms or signs of PE or suspected fetal growth restriction, GA 20^{+0} to 40^{+6} **Exclusion:** Confirmed PE diagnosis before enrollment	6

Table 2-6 Accuracy of PlGF tests in the prediction of adverse outcomes in HDPs

Author Year	Test/ Cut-off/ Assay	Outcome	Total (N) and Outcome (%)	Sensitivity (95%CI)	Specificity (95%CI)	LR+ (95%CI)	LR− (95%CI)	AUROC (95%CI)
A) MATERNAL OUTCOMES ONLY								
PlGF only								
Ghosh 2012	Serum PlGF <122 pg/ml (DRG)	PPH	766 (8.7%)	73.1 (60.7-82.9)	76.7 (73.3-79.7)	3.14 (2.57-3.82)	0.35 (0.24-0.52)	-
sFlt-1/PlGF ratio								
Leaños-Miranda 2013	Serum sFlt-1/PlGF ratio ≥871 (R&D Systems)	Composite	501 (9.5%)	52.1 (37.4-66.5)	77.9 (73.8-81.6)	2.36 (1.71-3.26)	0.61 (0.46-0.83)	-
Palomaki 2015	sFlt1/PlGF ratio; >85 **MOM** (Roche Diagnostics)	Composite	237 (8.9%)	61.9 (38.7–81.0)*	69.4 (62.8–75.4)*	2.0 (1.4–3.0)*	0.5 (0.3–1.0)*	-
Rana 2013*	sFlt1/PlGF ratio; ≥85 (Roche Diagnostics)	Composite	97 (8.2%)	100 (59.7–100)*	51.7 (40.9–62.3)*	2.1 (1.7–2.6)*	∞	-
B) PERINATAL OUTCOMES ONLY								
PlGF only								
Molvaerec 2013	Plasma PlGF; ≤12 pg/mL (Alere)	SGA	89 (24.7%)	72.7 (49.6-88.4)	62.7 (50.0-73.9)	1.95 (1.30-2.91)	0.44 (0.22-0.88)	-
sFlt-1/PlGF ratio								
Go´Mez-Arriaga 2014	sFlt-1/PlGF ratio >655	Composite neonatal	55 (27.5%)	92.8 (64.2-99.6)	54.1 (37.1-70.2)	2.02 (1.38-2.95)	0.13 (0.02-0.91)	0.75 (0.62–0.88)

Author Year	Test/ Cut-off/ Assay	Outcome	Total (N) and Outcome (%)	Sensitivity (95%CI)	Specificity (95%CI)	LR+ (95%CI)	LR− (95%CI)	AUROC (95%CI)
	ELISA (Roche Diagnostics)							
Leaños-Miranda 2013	Serum sFlt-1/PlGF ratio ≥871 (R&D Systems)	Stillbirths or neonatal deaths	501 (11.0%)	67.3 (53.2-79.0)	84.3 (80.5-87.5)	4.29 (3.23-5.69)	0.39 (0.27-0.57)	-
Leaños-Miranda 2013		small for gestational age infant (SGA)	501 (44.3%)	36.9 (30.6-43.7)	84.6 (79.7-88.5)	2.40 (1.73-3.31)	0.75 (0.67-0.83)	-
PlGF combined with other factors								
Go´Mez-Arriaga 2014	sFlt-1/PlGF ratio; >655 **+GA** (Roche Diagnostics)	Composite neonatal	55 (27.5%)	-	-	-	-	0.89 (0.79–0.99)
C) TIMED DELIVERY								
PlGF only								
Álvarez-Fernández 2016	Serum PlGF (Roche Diagnostics)	Delivery within the first week of clinical presentation **GA<34 weeks**	83 (25.3%)	-	-	-	-	0.89 (0.80–0.97)
Chaiworapongsa **2011** Chaiworapongsa **2014**	Plasma PlGF ≤0.4 **MOM** ELISA (R&D Systems)	Preterm delivery due to severe PE	87 (60.9%)	94.3 (84.6- 98.1)	70.6 (53.8- 83.2)	3.2 (1.9–5.4)	0.08 (0.03-0.25)	0.87 (0.79-0.95)
			85 (56.5%)	91.7 (79.1- 97.3)	62.2 (44.8- 77.1)	2.4 (1.6–3.7)	0.13 (0.05–0.4)	-
Chaiworapongsa **2011**	Plasma PlGF ≤0.15 **MOM**	Delivered within 2 weeks	59 (45.8%)	81.5 (63.3- 91.8)	84.4 (68.3- 93.1)	5.21 (2.29–12)	0.22 (0.10–0.49)	0.85 (0.75-0.95)

Author Year	**Test/ Cut-off/ Assay**	**Outcome**	**Total (N) and Outcome (%)**	**Sensitivity (95%CI)**	**Specificity (95%CI)**	**LR+ (95%CI)**	**LR− (95%CI)**	**AUROC (95%CI)**
Chaiworapongsa **2014**	ELISA (R&D Systems)	for **GA<34** weeks	43 (41.9%)	72.2 (46.4- 89.3)	92.0 (72.5- 98.6)	9.0 (2.3–35)	0.30 (0.1–0.6)	-
Chappell 2013	plasma PlGF; <5th centile for gestation (Alere Triage assay)	delivery for confirmed preeclampsia within 14 days. **GA<35 weeks**	287 (55.1%)	96.0 (89.0–99.0)	55.0 (48.0–61.0)	2.1 (1.8–2.5)	0.07 (0.02–0.22)	-
Molvaerec 2013	Plasma PlGF; ≤12 pg/mL (Alere)	Preterm delivery	89 (68.5%)	63.9 (50.6-75.5)	92.9 (75.0-98.7)	8.95 (2.32-34.48)	0.39 (0.28-0.55)	-
Ukah 2017	Serum PlGF; <100 pg/mL (Alere)	Preterm delivery	601 (18.0%)	28.0 (20.5-36.9)	89.4 (86.6-91.8)	2.66 (1.84-3.85)	0.80 (0.72-0.90)	-
Woelkers 2016	Serum PlGF; (Alere triage)	PE with Preterm delivery	753 (60%)	-	-	-	-	0.83 -
		PE with delivery within 14 days	753 (48%)	-	-	-	-	0.85 -
sFlt-1/PlGF ratio								
Chaiworapongsa **2011**	Plasma PlGF/sFlt-1; ≤0.005 **MOM**	Preterm delivery due to severe PE	87 (60.9%)	73.6 (60.4- 97.0)	91.2 (77.0- 99.0)	8.3 (2.8–25)	0.29 (0.18–0.46)	0.88 (0.81-0.96)
Chaiworapongsa **2014**			85 (56.5%)	66.7 (52.5- 78.3)	91.9 (78.7- 97.2)	8.2 (2.7–25)	0.36 (0.2–0.6)	-
Chaiworapongsa **2011**	Plasma PlGF/sFlt-1*; ≤0.035 **MOM**	delivered within 2 weeks **GA<34 weeks**	59 (45.8%)	92.6 (76.6- 97.9)	78.1 (61.3- 99.0)	4.23 (2.2–8.2)	0.09 (0.02-0.36)	0.88 (0.79-0.97)
Chaiworapongsa **2014**			43 (41.9%)	88.9 (63.9- 98.1)	96.0 (77.7- 99.8)	22.2 (3.23-152.69)	0.12 (0.03-0.42)	0.94

Author Year	**Test/ Cut-off/ Assay**	**Outcome**	**Total (N) and Outcome (%)**	**Sensitivity (95%CI)**	**Specificity (95%CI)**	**LR+ (95%CI)**	**LR− (95%CI)**	**AUROC (95%CI)**
De Oliveira 2013	Serum sFlt-1/PlGF ≥85 (Roche Diagnostics)	delivery due to severe PE	88 (46.5%)	74.4 (58.5- 86.0)	97.8 (86.8- 99.9)	33.5 (4.7- 234.4)	0.26 (0.16- 0.44)	0.954 (0.917–0.991)
Rana 2013	sFlt1/PlGF ratio; ≥85 (Roche Diagnostics)	Preterm delivery (<37)	80 (68.8%)	74.5 (60.7–84.9)*	96.0 (60.7–84.9)*	18.63 (2.71–127.6)*	0.37 (0.22-0.54)	-
PlGF combined with other angiogenic factors								
Chaiworapongsa **2011**	PlGF/ sEng; ≤0.07 **MOM** ELISA (R&D Systems)	Preterm delivery due to severe PE	87 (60.9%)	75.5 (62.4- 85.1)	91.2 (77.0- 99.0)	8.6 (2.9–25)	0.27 (0.17–0.44)	0.90 (0.83-0.97)
Chaiworapongsa **2014**			85 (56.5%)	66.7 (52.5- 78.3)	91.9 (78.7- 97.2)	8.2 (2.7–25)	0.36 (0.2–0.6)	-
Chaiworapongsa **2011**	PlGF/ sEng; ≤0.05 **MOM** ELISA (R&D Systems)	delivered within 2 weeks **GA<34 weeks**	59 (45.8%)	85.2 (67.5- 98.1)	84.4 (68.3– 93.1)	5.45 (2.4–12)	0.18 (0.07–0.44)	0.87 (0.77-0.96)
Chaiworapongsa **2014**			43 (41.9%)	88.9% (63.9- 98.1)	96.0 (77.7- 99.8)	22.2 (3.23 – 152.69)	0.12 (0.03-0.42)	0.94
D) COMBINED MATERNAL AND FETAL OUTCOMES								
PlGF only								
Rana 2012	Plasma PlGF; (Roche Diagnostics)	composite	80 (68.8%)					0.74 (0.70- 0.78)
sFlt-1/PlGF ratio								
Moore 2012	Serum sFlt-1/PlGF ratio (R&D Systems)	Composite	276 (28.3%)	-	-	-	-	0.76 (0.66-0.85)

Author Year	Test/ Cut-off/ Assay	Outcome	Total (N) and Outcome (%)	Sensitivity (95%CI)	Specificity (95%CI)	LR+ (95%CI)	LR− (95%CI)	AUROC (95%CI)
Rana 2012	Plasma sFlt1/PlGF ratio; ≥85 (Roche Diagnostics)	Composite 74.5 **Presenting <34 weeks**	176 (33.5%)	72.9 (59.5-83.3)	94.0 (87.6-97.4)	12.2 (5.8-25.4)	0.29 (0.19-0.44)	0.93 (0.89–0.97)
Rana 2012	Plasma sFlt1/PlGF ratio; ≥85 (Roche Diagnostics)	Composite at 2 weeks **Twins**	79 (65.8%)					0.75 (0.64–0.86)
		Composite at 2 weeks **Twins Presenting <34 weeks**	38 (57.9%)					0.81 (0.66–0.96)
PlGF combined with other factors								
Moore 2012	Serum sFlt-1/PlGF ratio (R&D Systems) + clinical multivariate model†	Composite **GA<37 weeks at presentation**	276 (28.3%)	-	-	-	-	0.91 (0.85-0.97)
Salahuddin 2016	Plasma sFlt1/PlGF ratio; ≥85 (KRYPTOR) **+ SBP + proteinuria**	Composite	412 (41.5%)	-	-	-	-	0.80 (0.76–0.85)
		Composite **GA<34 weeks at presentation**	110 (30.9%)	-	-	-	-	0.89 (0.82–0.95)

*Some zero cells

†Clinical multivariate model* (11 variables)**:** race, chronic hypertension, history of renal disease, gravidity (primigravid vs. multigravid), preeclampsia history, maternal age, smoking status, obesity (BMI >30 kg/m2), pre-gestational diabetes, clinical diagnosis of preeclampsia, and gestational age at presentation.

PE – Pre-eclampsia; **PPH** – Postpartum hemorrhage; **SGA** – small-for-gestational age

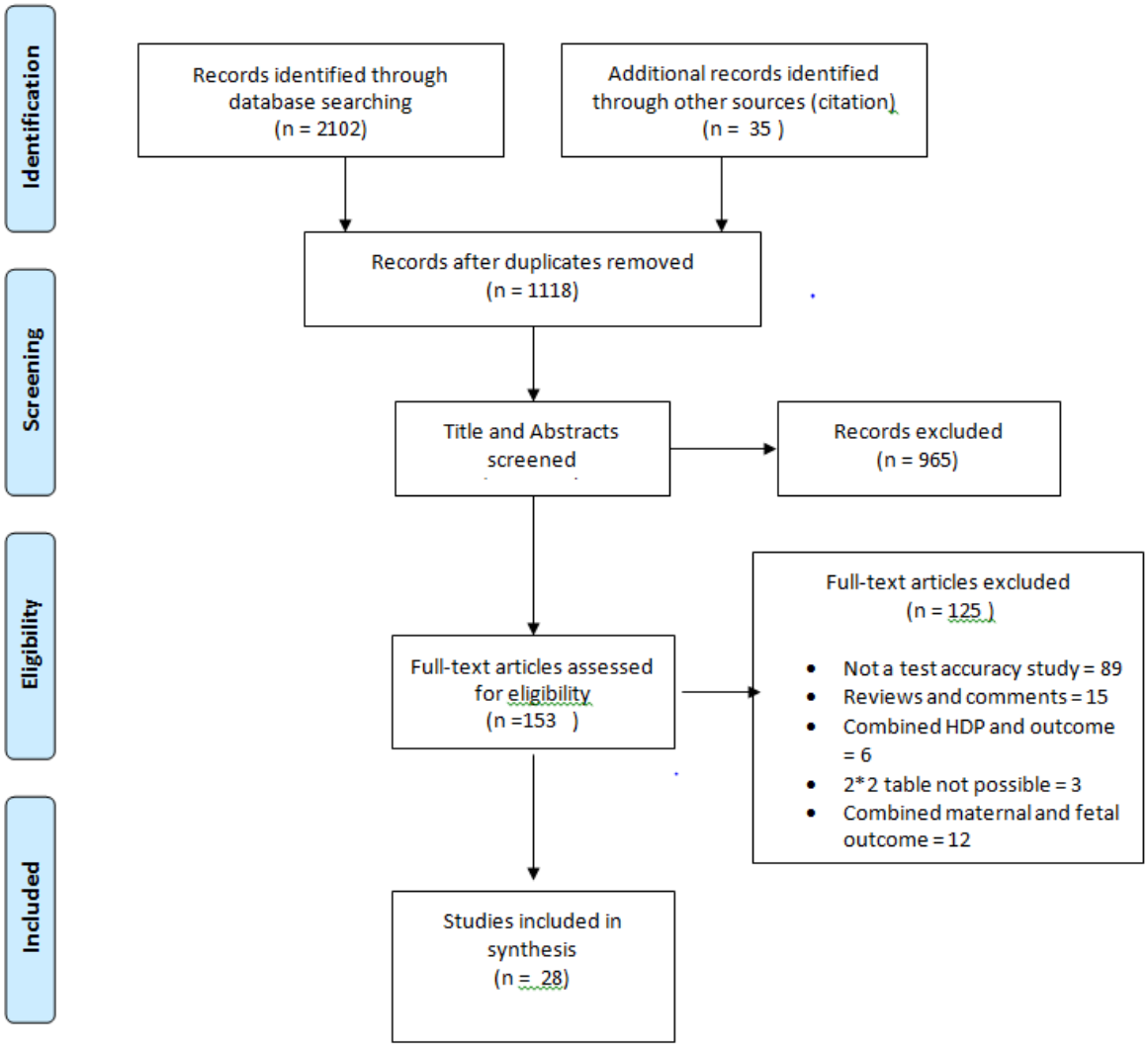


Figure 2-1 Article selection process

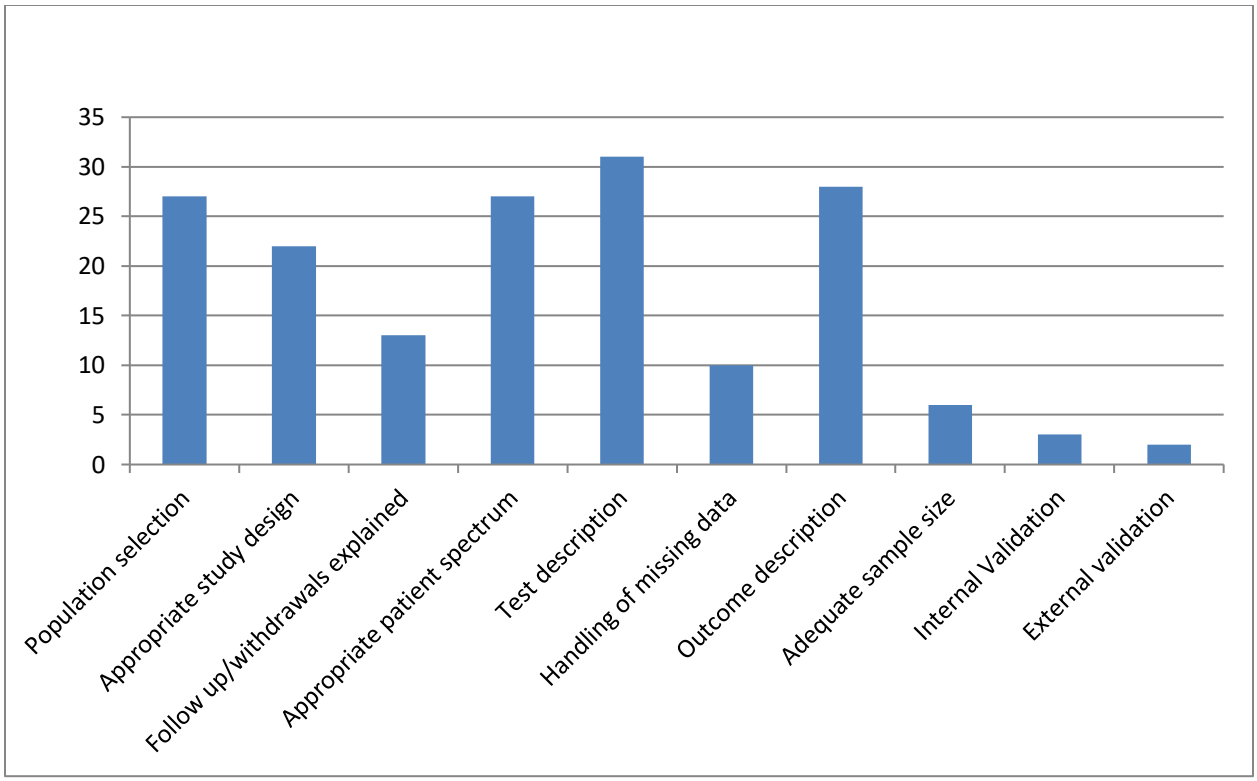


Figure 2-2 Quality of studies for Systematic review on prediction of maternal outcomes

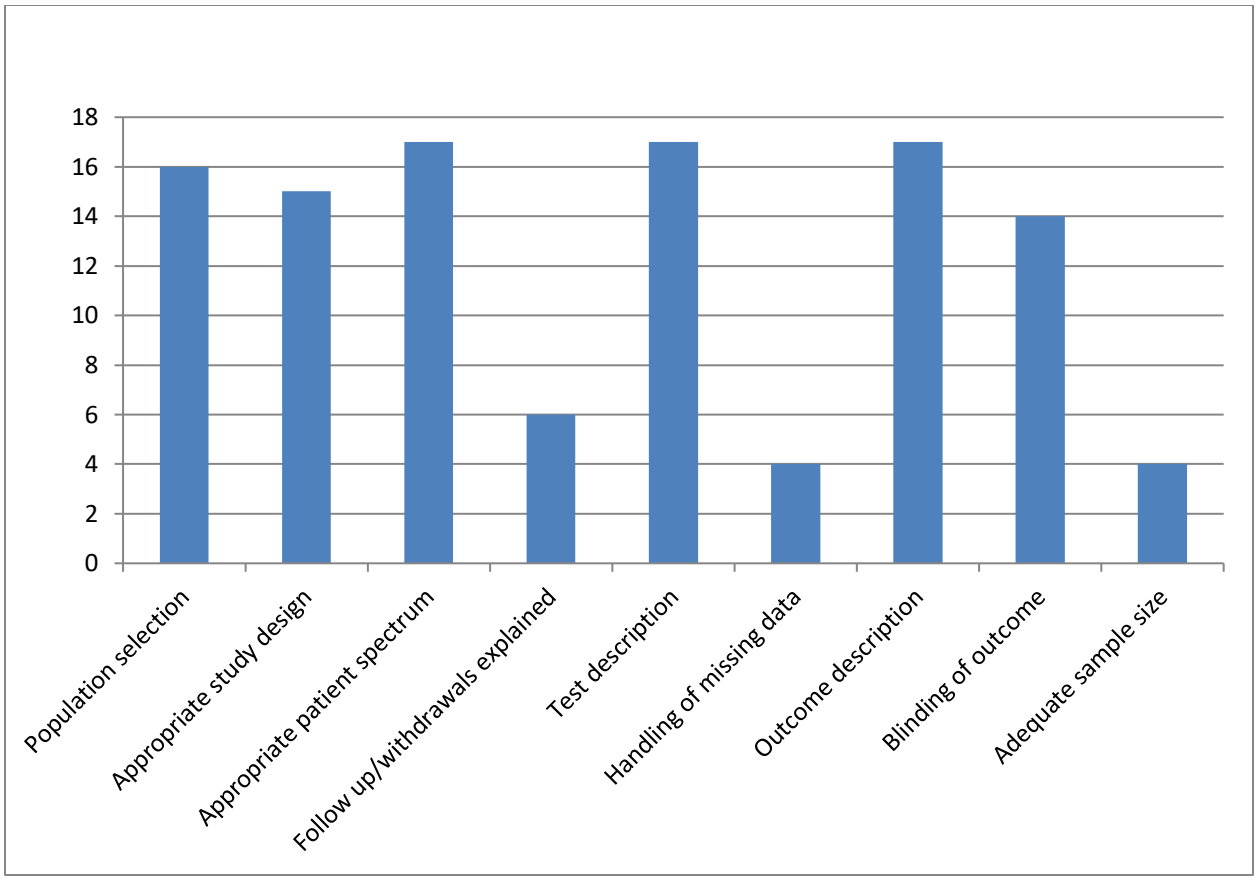


Figure 2-3 Quality of studies for Systematic review of PlGF

Chapter 3: Data collection procedures, missing data and model performance evaluation

This chapter is divided into three sections. Section 1 describes all the data used in this thesis, Section 2 discusses handling of missing data for the studies and Section 3 discusses methods used to evaluate the model performance in the studies.

3.1 Datasets

The cohorts used to meet the objectives of this thesis include datasets that were used for the development of the fullPIERS model, temporal and external validation, the assessment of the model in a broader disease spectrum and population, and extension of the model with the placental growth factor.

3.1.1 fullPIERS model development data

The fullPIERS model development data consisted of women prospectively collected from academic, level three (tertiary) obstetric centres in high-income countries.[24] The countries were Australia, Canada, New Zealand and the United Kingdom (UK). All the participating hospitals had expectant management policy for pre-eclampsia. In total, the data included 2,023 women admitted with pre-eclampsia from September, 2003 to January 2010. The rate of adverse outcome occurring within 48 hours of admission was 5%.

3.1.1.1 Inclusion and Exclusion criteria for the fullPIERS cohort

Inclusion criteria: Women admitted with pre-eclampsia defined as:

- blood pressure ≥140/90 mm Hg (at least one component, twice, ≥4 h apart, after 20 weeks) and either proteinuria (of ≥2+ by dipstick, ≥0·3 g per day by 24-h collection, or

≥30 mg/mmol by urinary protein:creatinine ratio) or hyperuricaemia (greater than local upper limit of local non-pregnancy normal range); (>345 μM);

- HELLP (haemolysis, elevated liver enzymes, low platelets) syndrome, even in the absence of hypertension or proteinuria; or
- superimposed pre-eclampsia (rapidly increasing requirements for antihypertensive drugs, systolic blood pressure >170 mm Hg or diastolic blood pressure >120 mm Hg, new proteinuria, or new hyperuricaemia).

Exclusion criteria:

Women were excluded from the cohort if they had already experienced an adverse maternal outcome before hospital admission or data collection or if they were admitted in spontaneous labour.

3.1.2 British Columbia Women's Hospital (BCW) data

The BC Women's Hospital cohort was made up of data extracted from the medical charts of 1,310 women admitted with pre-eclampsia between January 2012 and May 2016 at the BCW's hospital, Vancouver, Canada. This site was one of the centres that provided data for the development of the fullPIERS model. Medical records of women with documented HDPs were requested from the Hospital Health Information Analyst. These charts were reviewed and relevant information was extracted by trained data abstractors. To reduce data collection errors and misclassification, at least 5% of the charts were randomly reviewed again by a different abstractor other than the initial one the chart was assigned to. Data checks were also carried out during cleaning of the data and any errors or outliers were confirmed by going back to the charts. Any cases of uncertainty were resolved by discussion with the study coordinator (UVU) or the principal investigator (PvD). To reduce missing data, online medical records (PowerChart and

Cerner) were also used to retrieve more relevant information for cases. The same inclusion and exclusion criteria used in the development data collection were applied.

3.1.3 External validation cohorts

These data included cohorts that were collected outside any of the fullPIERS model development centres that were used to assess the validity of the fullPIERS model. **Table 3-1** presents the summary of datasets and their uses.

3.1.3.1 The Finnish Genetics of Preeclampsia Consortium (FINNPEC) cohort

The FINNPEC cohort was derived from data prospectively collected for the FINNPEC study[88,89] which aimed to identify genetic risk factors for pre-eclampsia. The dataset consisted of 1,450 women with singleton pregnancies diagnosed and admitted with pre-eclampsia at five university hospitals in Finland. Similar to the fullPIERS sites, these sites also practiced expectant management for pre-eclampsia. Women were excluded if they were less than 18 years old or could not provide consent for the FINNPEC study. For the purpose of this thesis, additional required fields for only women meeting the fullPIERS inclusion (the same definitions of pre-eclampsia) were requested from the study coordinator. Information such as laboratory measurements and adverse maternal outcomes, were abstracted from hospital medical records of the women by trained research assistants/medical students in the centres. Any data concerns were resolved by discussion and revisiting the medical charts. In total, the cohort which met the fullPIERS inclusion criteria included 124 women admitted between January 2008 and May 2011.

3.1.3.2 Pre-EcLampsIa: Clinical ApplicatioN (PELICAN) cohort

The PELICAN cohort was derived from data prospectively collected for the Alere-funded PELICAN study.[83] The aim of this study was to assess the diagnostic accuracy of PlGF for pre-eclampsia requiring delivery within two weeks of sample testing. The data consisted of 625

women who presented with suspected symptoms and signs of pre-eclampsia between January 2011 and February 2012 at seven consultant-led maternity units in the UK and Ireland; women were excluded if they were less than 16 years old. These sites are also known to practice expectant management for women with pre-eclampsia. Similar to the FINNPEC data, additional required fields were requested from the study site coordinators for only women with confirmed pre-eclampsia as defined in the fullPIERS study and data were abstracted from hospital medical records by trained research assistants/medical students. In total, the cohort included 70 women admitted into one of the study sites (St. Thomas Hospital, London) were included; the PlGF measurements for these women were also recorded.

3.1.3.3 Pre-Eclampsia Triage by Rapid Assay (PETRA)

The Alere-PETRA cohort was derived from the Alere-funded PETRA prospective, observational study[86] which, similar to the PELICAN cohort, aimed at assessing the diagnostic accuracy of PlGF and also to predict women with requiring delivery within two weeks of sample testing due to pre-eclampsia. The PETRA data consisted of 1,223 women with suspected symptoms and signs of pre-eclampsia presenting at 24 maternity units in the United States of America (USA) (N=23 units) and Canada between May 2010 and November 2011. Majority of the study centres were known to practice interventionist management for pre-eclampsia. In total, the cohort included 644 women meeting the fullPIERS inclusion criteria. The PlGF measurements for these women were also recorded.

3.1.3.4 John Radcliffe Hospital (JRH) cohort (Oxford)

The JRH cohort data were made up of data extracted from the medical flow sheets of 281 women admitted with pre-eclampsia between January 2003 and December 2006, at the Silver Star Unit (tertiary maternity unit) of the John Radcliffe Hospital, Oxford, UK. Missing information was

obtained from the electronic system, Patient administration system (PAS) and any data uncertainties were resolved by discussion with the centre coordinating investigator and an obstetric physician.

The same inclusion and exclusion criteria used in the development data collection were applied and management was expectant.

3.1.3.5 The Pre-eclampsia Eclampsia TRial Amsterdam (Dutch PETRA) cohort

The Dutch PETRA cohort was obtained from the data collected for the Dutch PETRA randomized controlled trial which investigated the effect of plasma volume expansion on the pulsalitility indices of the fetal cerebral arteries. The cohort consisted women admitted between 24 to 34 weeks of gestation with severe pre-eclampsia (defined as pre-eclampsia with diastolic blood pressure ≥110mmHg), HELLP syndrome, eclampsia or hypertension-related fetal growth restriction at two tertiary care obstetric hospitals in the Netherlands. Women were excluded if they had signs of fetal distress or maternal disease demanding immediate delivery, or a pre-existing diagnosis of a lethal fetal congenital abnormality. In total, the cohort included 216 women admitted between April 2000 and May 2003.

3.1.3.6 Prediction of complications in early-onset pre-eclampsia (PREP)

The PREP cohort consists of women prospectively collected from secondary and tertiary obstetric units in the UK.[91] The cohort included 1,096 women admitted to a participating hospital unit between December 2011 and April 2014 with suspected pre-eclampsia before 34 weeks of gestation (early-onset pre-eclampsia). The inclusion and exclusion criteria used were the same as the fullPIERS study. In total, the data included 946 women admitted with pre-eclampsia who met the fullPIERS eligibility criteria.

3.1.3.7 miniPIERS

This cohort consists of women prospectively collected from level two (secondary) and level three (tertiary) obstetric centres in low-resourced settings.[62] The five LMICs were Fiji, Uganda, South Africa, Brazil, and Pakistan. The cohort included women admitted to a participating hospital unit with any HDP i.e. pre-eclampsia similarly defined as in the fullPIERS study, or gestational hypertension defined as blood pressure (BP) ≥140/90 mmHg (at least one component, twice, ≥4 hours apart, ≥20+0 weeks) without significant proteinuria, or chronic hypertension as BP ≥140/90 mmHg (at least one component, twice, ≥4 hours apart, <20+0 weeks' gestation) or full or partial partial HELLP (i.e., haemolysis and low platelets OR low platelets and elevated liver enzymes). In total, the data included 2,081 women admitted from July 2008 to March 2012.

3.2 Missing Data

Missing data is a problem in validation studies due to the use of secondary data.[92] Appropriate handling of missing data is necessary to avoid often misleading and biased results which could result from selection bias and lack of power.[92-94] Imputation of missing data has been proposed as better alternative to using only complete data (complete-case analyses).[93,94] Multiple imputation is the preferred method as it generates several plausible values for the missing variable values rather than imputing missing values with data means or medians, which could skew the distribution of the variable.[92-95] This section describes handling of missing data within the cohorts.

3.2.1 Classification of missing data (mechanism)

Before carrying out any analyses for cohorts with missing data, cases with complete data are compared to cases with incomplete data to explore the pattern of missingness.[89] Based on clinical knowledge of the data collection sites, certain variables were already considered missing

completely at random (MCAR) in some centres where the tests, such as SpO_2, are not used. Data are assumed to be 'Missing at random' (MAR) if the pattern of missingness can be explained by variable(s) in the data and assumed to be 'Missing Not at random' (MNAR) if there are no explanations. For MAR and MCAR,[92] multiple imputation which has been reported to work optimally for these patterns of missingness, was used in this study as described below.

3.2.2 Multiple imputation

Multiple imputation analyses were carried out to generate plausible values using the Multiple Imputation by Chained Equations (MICE) package in R program.[94] Predictive mean matching based on Rubin's formula was used to generate the missing values for numeric variables such as AST, platelet and creatinine while logistic regression was used for binary variables such as chest pain. This method of imputation produces possible values using chained equation models based on specified variables. It is recommended to include as many variables as possible, including the study outcomes in the imputation models, to get more reliable values.[92,94] Therefore for our study, we included all the model variables and outcomes as well as auxiliary variables which included maternal age, parity, other measured symptoms and laboratory values.

This imputation technique was carried out ten times resulting in ten datasets. The predicted risk of developing an adverse outcome was calculated in each of the ten derived datasets using the published fullPIERS equation. The final predicted risks were combined by averaging the ten calculated predicted risks for each woman with pre-eclampsia.[96] This average predicted probability was then used to evaluate the predictive performances of the model (in the imputed dataset), described below.

3.3 Model performance evaluation

This section describes the methods used to evaluate the performance of the fullPIERS model in the study. Model evaluations used include traditional methods such as discrimination, calibration and stratification as well as new methods such as the Net Reclassification Index (NRI) and integrated discrimination improvement (IDI).

3.3.1 Discrimination

Discrimination measures how accurate the model is in distinguishing between those with or without adverse outcomes. Discrimination was summarized by the concordance (c) index also known as the area under the receiver-operator characteristic curve (AUROC) by plotting sensitivity/true positive rate (on the Y-axis) against the false positive rate (on the X-axis).[21,97] Sensitivity for this study was defined as the percentage of correctly predicted adverse maternal outcomes as compared with all observed adverse maternal outcomes, whereas specificity was defined as the percentage of correctly predicted women without adverse maternal outcomes as compared with all women without adverse maternal outcomes. The AUROC was interpreted using the following pre-specified criteria: non-informative (AUROC $\leq 0{\cdot}5$), poor discrimination ($0.5 <$ AUROC $\leq 0{\cdot}7$), good discrimination (AUROC $> 0{\cdot}7$).

3.3.2 Calibration

Calibration was measured by plotting the mean predicted risks of adverse outcomes on the X-axis and the observed outcomes on the Y-axis. [21,97] The curve was compared to a curve of ideal calibration curve represented by fitting a 45-degree straight line on the calibration graph, and then, the calibration slope will be estimated. In the case of a perfect calibration, all the predictions and observations would be on the ideal calibration line (slope will be 1 and intercept will be 0). However, due to study population differences in external validation, the calibration is

almost never perfect. 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. 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 (1.30 < slope > 0·7).

3.3.3 Stratification

Stratification measures the proportions in which subjects with pre-eclampsia are assigned to the different clinically relevant risk categories i.e. classification of women into low- and high- risk categories using a classification table with generated risk groups.[100] A predicted probability score of ≥0.30 was used for high risk category and <0.025 for low risk category in the fullPIERS model development. The fullPIERS model stratification capacity was assessed based on the extent to which it assigns the women who develop the combined adverse outcomes into high risk category and assigns women who do not develop adverse outcomes to the low risk category in the validation datasets, at these prediction scores.

The Likelihood Ratios (LRs) using Deek and Altman's method,[34] true and false positive rates, negative predictive values (NPVs), and positive predictive values (PPVs) were also computed for each risk stratification group to show how well the model performs within each group instead of the overall model.

3.3.4 Net reclassification index (NRI)

NRI measures the number of women that were reclassified correctly into low and high risk groups in the extended model with PlGF compared to the original fullPIERS model.[21,97,101] We used the same threshold in the fullPIERS model development (predicted probability score of ≥0.30 was categorised as highest risk group and a predicted probability score of <0.025 was

categorised as low risk group) or another threshold depending on the distribution of the calculated predicted risks. NRI gives a summary of the overall improvement of sensitivity and specificity by integrating the "upward movement" (improved reclassification) and the "downward movement" (worse reclassification) for the women with adverse outcomes in the predicted risk groups with the reverse movements for the women without adverse outcomes.. Any increases in the highest risk group with correctly identified cases of adverse outcomes and in the low risk group with correctly identified cases without adverse outcomes implied improvement in classification with the extended model (that is, the overall increase in sensitivity and specificity within the threshold). The sum of the differences between the two movements for the women with and without adverse outcomes was the NRI.

3.3.5 Integrated discrimination improvement (IDI)

IDI is the difference in the discrimination slope between the original model and the extended model.[8,97,101] The discrimination slope is calculated as the difference between the average predicted probabilities of the women with and without adverse maternal outcomes.

3.4 Statistical analyses

All statistical analyses were performed using R version 3·1·3 (The R Project for Statistical Computing) and Microsoft Excel program (version 3, Redmond, WA).

For any hypothesis test, a p-value of < 0.05 was considered significant.

Table 3-1 Summary of Datasets

Datasets	Centres of enrollment	Thesis Use
fullPIERS	High income Countries: Canada, Australia, United Kingdom and New Zealand	Development data
BCW	High income Countries: Canada	Temporal and Primary External validation
FINNPEC	High income Countries: Finland	Primary External validation
PELICAN	High income Country: United Kingdom	Primary External validation
JRH (Oxford)	High income Country: United Kingdom	Secondary External validation
Alere-PETRA	High income Countries: USA and Canada	Secondary External validation
Dutch-PETRA	High income Country: The Netherlands	Broader spectrum assessment
PREP	High income Country: United Kingdom	Broader spectrum assessment
miniPIERS	Low-and-middle-income countries: Pakistan, Fiji, South Africa, Uganda, Brazil	Broader spectrum assessment

Chapter 4: External (geographical and temporal) validation of the fullPIERS model

4.1 Background

As stated earlier, the fullPIERS model has a promising predictive performance. However, temporal and external validations of the model are necessary before it should be implemented for clinical use.[21] Validation ascertains the accuracy of the model in the population that it was designed for based on discriminatory and calibration performance of the model.[22,23] Temporal validation is carried out in the same setting as the one used in the model development but with more recent patients, thereby prospectively evaluating the model.[21-23] External validation involves assessments 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.[21-23]

Although, there have been a few attempts to externally validate the fullPIERS model, none of the studies conducted have used similar populations and settings as the development population. Differences include use of data from low-and-middle-income countries (LMIC)[98] or using a subset (less than 34 weeks gestational age)[90] or broader inclusion criteria for disease (all HDPs).[99] In addition, inadequate sample size has been an issue limiting the conclusions made from these studies. Therefore, an assessment of external validity of the fullPIERS model using datasets collected in high income countries with similar populations was conducted.

4.2 Methods

4.2.1 Temporal and external validation datasets

Five of the 8 cohorts described in **Chapter 2** were deemed appropriate for the temporal and external validation of the fullPIERS model, based on having similar inclusion and exclusion criteria, and admission into tertiary units in high-income settings. The data included were collected from the (i) BCW Hospital (ii) the FINNPEC study (iii) PELICAN study (iv) Alere-funded PETRA study, and (v) Oxford data.

Due to differences in clinical practice (Alere-PETRA data) and data completion (Oxford data), the external validation was divided into two sections:– primary external validation (to assess the reproducibility of the model in the most similar and reliable datasets) and "broad" external validation (to assess the overall transportability/ generalizability of the model). The *a priori* plan was to base the interpretation of the model performance on the primary external dataset results. In addition, the model was assessed in only the BCW cohort for temporal validation because it was one of the model development sites. Therefore, there were 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.

4.2.2 Definition of pre-eclampsia and outcomes

The same definitions for pre-eclampsia and adverse outcomes used in the model development,[24] as described in **Chapter 3**, were used.

The same primary outcome in our validation study was the same as in the model development study. When the exact time of day of the occurrence of outcome was unknown, outcomes occurring within two calendar days from the date of admission were used as a proxy.

4.2.3 Statistical analyses

4.2.3.1 Demographics

The distribution of patient characteristics in each of the three validation sets were compared with the development (fullPIERS) cohort.[24] Univariate comparisons of the characteristics of women who experienced an adverse outcome and those who did not were also performed for each validation data set.

4.2.3.2 Model performance evaluation

Using the worst value of each predictor measured within 48 hours of admission and prior to the occurrence of an adverse outcome, the fullPIERS equation was applied to 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 as described in **Chapter 3**. Likelihood Ratios (LRs), true and false positive rates, NPVs, and PPVs were also computed for each risk stratification group.

4.2.3.3 Missing data

Missing SpO_2 was substituted with 97%, as was done during the fullPIERS model development. If missing, AST was substituted with alanine transaminase (ALT) when available, as this measurement had been agreed to be biologically similar during the model development. 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 imputation was used to generate plausible values for missing variables. Multiple imputation was

carried out ten times using the multiple imputation by chained equations (MICE) method in R statistical program (The R Project for Statistical Computing).

For sensitivity analyses, we carried out imputation twenty times for the combined validation data due to the higher rate of missing data. We also examined the model performance imputing SpO_2 values using multiple imputation rather than the median of 97%.

4.2.3.4 Sensitivity analyses

Secondary analyses included assessment of the discrimination capacity of the (i) primary and (ii) combined external validation data without the BCW cohort. The performance of the model for the prediction of adverse maternal outcomes within 7 days was also assessed.

4.2.3.5 Sample size for External Validation

Our sample size was guided by simulation studies, which recommend that validation studies should have at least 80 to 100 events (outcomes) in order to have 80% power at the 5% significance level.[102]

4.3 Results

4.3.1.1 Comparison of the development and validation cohorts

In total, the combined cohort included 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 4-1**. 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 less likely to be multiparous and more likely to have symptoms of chest pain or dyspnoea, while the PELICAN cohort had a lower rate of smoking, and higher uric acid measurements.

The women in the PETRA cohort were more likely to have an earlier onset of pre-eclampsia, had the highest 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 were more likely to have higher platelets and creatinine measurements.

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 4-2**. In total, the primary external, temporal validation and broader cohorts included 1504, 1310, and 2429 women respectively.

Within 48 hours of admission, the rates of adverse maternal outcomes experienced 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.

4.3.1.2 Adverse maternal outcomes

The most common outcomes in the combined validation cohorts were blood transfusion (N=61), placental abruption (N=21), and infusion of a third antihypertensive medication (N=21) **(Table 4-3)**. There were no cases of maternal deaths, cortical blindness or hepatic rupture.

In all the validation cohorts, women with an adverse maternal outcome within 48 hours had an earlier onset of pre-eclampsia, appeared to have worse clinical measures (e.g. higher chest pain, sBP, uric acid, and lower platelet count) and more interventions (antihypertensive and magnesium sulphate treatment) (**Table 4-4**). They also delivered at an earlier gestational age to babies of lower birth weights.

4.3.1.3 Data completeness

Gestational age was the most complete variables 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 cases of missing data (**Table 4-5**). The broader validation dataset had the most missingness: 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.

The comparison between women with complete predictor variables data and women with missing data in the primary external data showed that the women with missing data were more likely to be younger, multiparous and appeared to have higher platelet counts, lower creatinine and AST (

Table 4-6).

4.3.2 Model performance

4.3.2.1 Primary external validation

The fullPIERS model showed good discrimination in the primary external validation datasets with an AUROC of 0.81 (95% CI: 0.76-0.87) (**Figure 4-1a**). 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 a good calibration performance in the primary external validation dataset with a slope of 0.70 although the intercept was marginally elevated ($\alpha = 0.3$) (**Figure 4-2**).

Table 4-7 shows the distribution of women in various probabilities risk groups, similarly used in the model development, and the stratification performance of the model in each group for the primary external validation dataset. The distribution of women in each group was similar to the development cohort with 30.5% of women having a predicted probability of <1% and 4% of women with a predicted probability of ≥30%. 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%).

4.3.2.2 Temporal validation

For temporal validation using only the BCW cohort, the fullPIERS model also showed good discrimination capacity with an AUROC of 0.82 (95% CI: 0.76-0.87) (**Figure 4-1**c), which did not change after imputation of missing values. Calibration was good with a slope of 0.70 and intercept of 0.20 (**Figure 4-2**b).

Table 4-8 presents the distribution of women and the stratification performance of the model in the temporal validation datasets in each predicted risk groups. The proportions of women in 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 was (56%) was also similar to the proportion in the development data (59%). The resulting false positive rate was also 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.

4.3.2.3 Broader validation

The fullPIERS model retained a good discrimination in the broader validation dataset although the AUROC decreased (0.74 (95% CI: 0.69-0.80)) (**Figure 4-1**c). There was also no significant change after imputation (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 (**Figure 4-2**).

Similar to the primary external dataset, about 4% of women had a predicted probability of ≥30% (Table 4-9). At this highest risk group, half of the women had an adverse outcome. The resulting false positive rate was 2% (specificity 98%) and the true positive rate was 27% with a high 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.

4.3.3 Sensitivity analyses

4.3.3.1 Performance for outcomes within seven days

The model performance decreased in all the datasets for the prediction of adverse outcomes within seven days. The AUROCs for the datasets after imputation were 0.71 (95% CI: 0.66-0.76), 0.78 (95% CI: 0.72-0.83), and 0.69 (95% CI: 0.65-0.73), for the primary external, temporal, and broader validation datasets, respectively. Calibration was poor in all datasets and decreased from temporal to primary to broader validation datasets with slopes of 0.63, 0.61 and 0.44, respectively.

4.3.3.2 Performance of model excluding the BCW cohort

The discriminatory performance of the model dropped in the primary and broader external validation datasets upon the exclusion of the BCW's cohort compared to when the BCW cohort was included, although the AUROCs remained >0.70 (**Figure 4-3**).

4.3.3.3 Imputation of SpO_2

Imputation of missing data including SpO_2 showed a trend of increasing discriminatory performance compared to using the median of 97% in all the validation datasets, although the observed increases were not significant. The resulting AUROCs were 0.81 (95% CI: 0.76-0.87), 0.83 (95% CI: 0.78-0.88), and 0.78 (95% CI: 0.74-0.82), for the primary external, temporal and broader validation datasets, respectively.

4.4 Discussion

4.4.1 Main findings

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 the validation cohorts. The model was able to classify women into low and high-risk groups using a predicted probability cut-off of ≥30%; it also showed a strong ability to 'rule in' adverse outcomes within 48 hours at this cut-off.

All the AUROCs in the validation cohorts were lower than the original fullPIERS model performance of 0.88 (95% 0.84-0.92),[21] although only significantly lower in the broader validation cohort (the 95% confidence intervals overlapped for the primary external and temporal validation). The decrease in the model performance could be caused by overfitting in the original model development, case-mix differences, or differences in the effects of the model predictor

between the development and validation cohorts, or a combination of some or all.[21,103,104] This decrease in performance was also observed in the calibration ability upon temporal validation, whereby the most similar dataset compared to the development cohort was used. There were few case-mix differences between the development and temporal validation data with regard to demographics, antihypertensive administration, and adverse neonatal and maternal outcomes **(Table 4-1)**. 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.[103,104] We suspect that it is more likely that these reduced performances were due to different predictor effects and, perhaps, overfitting rather than due to case-mix differences. Case-mix differences may have played a more substantial role in the broader external dataset, in addition to the predictor-effects, as there were more differences in population characteristics more observed in the PETRA and Oxford datasets compared to the fullPIERS cohort. There was also a slightly higher rate of adverse maternal outcome within 48 hours, which was reflected in the calibration intercept (0.3).

Of note, the PETRA cohort[90] which significantly contributed to the broad validation dataset primarily included women admitted in the USA. The pattern of practice for the management of women with pre-eclampsia in the USA is different from the other datasets in that the women in the cohort were less likely to be managed expectantly. This could also be observed from the short admission to delivery interval for women with gestational age less than 34 weeks compared with the other cohorts **(Table 4-1)**. In addition, this cohort contributed most to the use of a third IV antihypertensive, an outcome which was rare in the other cohorts. This suggests either a different pattern of management for women with pre-eclampsia or that the women in this

cohort were more likely to have resistant hypertension compared to the other cohorts. 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 or differences in predictor effect.[103,104] These findings suggest a need for recalibration of the model to improve its performance if it is to be used in broader patient populations.[21]

Multiple imputation of SpO_2 showed a general trend of increasing performance of the model in the datasets compared to using the median, although this was not significantly different. When using the model in settings with some missing values for SpO_2 as well as other variables, multiple imputations can be used. However, for settings where there are no measurements at all (missing at random [MAR]), such as from the Oxford cohort, then the median of 97% can be used since one can only impute a variable when there are some measured values for the same variable in the dataset. It is in hope that these results will encourage the measurements of SpO_2 and other variables in the fullPIERS model since they are important predictors of adverse maternal outcomes within 48 hours.

4.4.2 Strengths and limitations

A strength of this study is the assessment of the validity and transportability of the fullPIERS risk prediction model in 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.[21,102] 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 we included women from BCW's Hospital 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 different from those used in the model development and, therefore, can still be considered as an external cohort, even though the study site is the same. We also assessed the discriminatory performance of the datasets without the BCW cohort and the AUROC remained >0.70 showing that the model still had good discriminatory performance and its performance was not entirely dependent on the BC Women's Hospital cohort. Another limitation was the large percentage of missing data in the broader external dataset. Although we accounted for these by imputation, these may have affected the model performance point estimate and confidence intervals. However, research suggests that imputation is preferable to omission of individuals, even if a predictor is completely missing in a dataset.[105,106,107] Our imputation analyses did not show any significant difference, thereby suggesting that there was less likelihood of bias in the broader external datasets.

4.4.3 Interpretation

Our external validation showed that the fullPIERS model is useful in discriminating between patients at high and low risk of adverse maternal outcomes within 48 hours and even up to a week in the temporal cohort. Our study also shows that using a threshold of ≥30% predicted risk was a good threshold to rule-in the outcome. Although we used this threshold to be consistent with the development study, it might be useful to explore different threshold for better accuracy. Based on our results, the model can be used to aid clinicians in managing women with pre-eclampsia in similar settings and 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 conservative management of pre-eclampsia such as the PETRA,

since there was a significantly reduced performance. Recalibration of the model should be considered in these settings before clinical use.

4.5 Conclusion

The fullPIERS model is temporally and externally valid for the prediction of adverse maternal outcomes occurring within 48 hours of admission for pre-eclampsia. Recalibration might be helpful in improving the calibration performance for settings with different patterns of clinical management. Future studies should explore recalibration and assessing the model performance in different sub-groups using fully powered data.

Table 4-1 Distributions of maternal characteristics for the individual validation datasets

Characteristics	fullPIERS cohort (development) (N=2,023)	BCW *(Also Temporal)* (N=1,310)	FINNPEC (N=124)	PELICAN (N=70)	PETRA (N=644)	Oxford (N=281)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS						
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	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	67 [58, 77]	64 [56, 75]	62 [54, 69]	70 [59, 84]	61 [53, 71]	75 [68, 84]

Characteristics	fullPIERS cohort (development) (N=2,023)	BCW *(Also Temporal)* (N=1,310)	FINNPEC (N=124)	PELICAN (N=70)	PETRA (N=644)	Oxford (N=281)
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	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 48h	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*

Table 4-2 Maternal characteristics for the grouped validation datasets

Characteristics	fullPIERS cohort (development) (2,023 women)	Temporal± (1 cohort) (1,310 women)	Primary external± (3 cohorts) (1,504 women)	Broader± (5 cohorts) (2,429 women)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS				
Maternal age at EDD (yr)	31 [27, 36]	34 [31, 38]	34 [30, 38]	33 [29, 37]
Parity ≥1	581 (28·7%)	409 (31.2%)	465 (30.9%)	872 (35.9%)
Gestational age at eligibility (wk)**	36 [33, 38·3]	37.7 [35.6, 39]	37.4 [35, 38·9]	36.6 [33.0, 38.4]
Gestational age at eligibility <34 weeks, N	636 (31.4%)	218 (16.6%)	285 (19.0%)	715 (29.4%)
Multiple pregnancy	192 (9·5%)	136 (10.4%)	141 (9.4%)	219 (9.0%)
Smoking in this pregnancy	249 (12·3%)	90 (6.9%)	107 (7.1%)	270 (11.1%)
CLINICAL MEASURES				
Systolic BP (mm Hg)	160 [150, 176]	160 [151, 171]	160 [151, 171]	156 [145, 168]
Diastolic BP (mm Hg)	102 [98, 110]	100 [94, 105]	100 [94, 106]	96 [89, 103]
Chest pain/dyspnoea**	90 (4.4%)	82 (6.3%)	95 (6.3%)	116 (4.8%)
Uric acid	376 [320, 427]	379 [323, 436]	379 [323, 436]	371 [313, 428]
Lowest platelet count ($\times 10^9$ per L)**	192 [150, 242]	174 [136, 217]	176 [136, 219]	188 [145, 233]
Highest AST/ALT (U/L)**	28 [21, 41]	33 [26, 47]	32 [24, 46]	28 [21, 42]
Creatinine	67 [58, 77]	64 [56, 75]	64 [56, 76]	64 [55, 76]
INTERVENTIONS DURING ADMISSION				
Corticosteroids	550 (27·2%)	320 (24.4%)	407 (27.1%)	646 (26.6%)
Corticosteroids, GA onset <34	440/636 (69.2%)	195/218 (89.5%)	252/285 (88.4%)	454 (18.7%)
Antihypertensive therapy	1381 (68·3%)	896 (68.4%)	1058 (70.4%)	1696 (69.8%)
$MgSO_4$	690 (34·1%)	393 (30.0%)	473 (31.5%)	968 (39.9%)
PREGNANCY OUTCOMES				
Admission-To-Delivery Interval (Days)	2 [1, 5]	1 [1, 3]	2 [1, 4]	2 [1, 5]

Characteristics	fullPIERS cohort (development) (2,023 women)	Temporal± (1 cohort) (1,310 women)	Primary external± (3 cohorts) (1,504 women)	Broader± (5 cohorts) (2,429 women)
Admission-To-Delivery Interval, <34^{+0} Weeks (Days)	4 [2, 14]	4 [2, 11]	6 [2, 11]	4 [2, 11]
Gestational age at delivery (wk)	36.9 [34·1, 38·6]	37.8 [36, 39.1]	37.7 [35.9, 39]	37 [34, 38.6]
Birth weight	2141 [1441, 2807]	2885 [2275, 3364]	2825 [2185, 3330]	2640 [1810, 3225]
Stillbirth	20 (1.0%)	7 (0.5%)	7 (0.5%)	25 (1.0%)
Neonatal death	26 (1·3%)	10 (0.8%)	11 (0.7%)	25 (1.0%)
MATERNAL OUTCOME (N women)				
Within 48h	106 (5.2%)	87 (6.6%)	99 (6.7%)	171 (7.0%)
Within 7 days	203 (10.0%)	110 (8.4%)	152 (10.1%)	253 (10.42%)
At anytime	261 (12.9%)	122 (9.3%)	190 (12.6%)	313 (12.3%)

± Temporal – BCW; Primary External – BCW, FINNPEC and PELICAN; Broader – Combined- BCW, FINNPEC and PELICAN, Alere PETRA and Oxford

***Variables included in the model*

Table 4-3 Adverse maternal outcomes in Validation datasets within 48 h for women

Events	Temporal (BCW) N women = 87	Primary External N women = 99	Broad External N women = 171
Maternal death	0	0	0
CNS	**3**	**9**	**14**
Eclampsia	2	7	10
Glasgow Coma <13	1	2	2
Stroke	0	0	1
TIA	0	0	0
Cortical blindness	0	0	0
Posterior reversible encephalopathy	0	0	1
Cardiorespiratory	**25**	**29**	**60**
Positive ionotropic support	2	2	2
Infusion of a third parenteral antihypertensive drug	1	2	21
Myocardial ischaemia or infarction	0	0	2
Require O_2 ≥50% for >1h	11	13	13
Intubation	4	5	7
Pulmonary oedema	7	7	15
Haematological	**59**	**65**	**91**
Transfusion (where recorded)	39	45	61
Platelet count < 50	20	20	30
Hepatic	**3**	**3**	**3**
Hepatic dysfunction	3	3	3
Hepatic Rupture	0	0	0
Renal	**11**	**11**	**19**
Creatinine > 150 without renal disease/ Creatinine > 200 with renal disease	11	11	19
Placental abruption	**13**	**15**	**21**
Ascites	**1**	**1**	**1**
Bell's Palsy	**2**	**2**	**3**
Hysterectomy/ruptured uterus	**0**	**0**	**1**
Any maternal event	**117**	**135**	**213**

Table 4-4 Maternal characteristics for the Temporal (N= 1310), Primary external (N= 1504), Broader (combined) datasets (N= 2429)

Characteristics	Temporal		Primary external		Broader	
	Outcome (87 women)	No outcome (1223 women)	Outcome (99 women)	No outcome (1405 women)	Outcome (171 women)	No outcome (2258 women)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS						
Maternal age at EDD (yr)	35 [31, 38]	34 [31, 38]	35 [31, 38]	34 [30, 38]	33 [29, 37]	33 [28.8, 37]
Parity ≥1	23 (26.4%)	386 (31.6%)	29 (29.3%)	436 (31.0%)	59 (34.5%)	813 (36.0%)
Gestational age at eligibility (wk)**	36.4 [33.9, 38.7]	37.7 [35.7, 39.0]	36.4 [33.9, 38.7]	37.4 [35.1, 39.0]	35.6 [32.8, 38.1]	36.7 [33.1, 38.4]
Gestational age at eligibility <34 weeks, N	23 (26.4%)	195 (15.9%)	26 (%)	259 (18.4%)	62 (36.3%)	653 (28.9%)
Multiple pregnancy	14 (16.1%)	122 (10.0%)	14 (26.3%)	127 (9.0%)	22 (12.9%)	197 (8.7%)
Smoking in this pregnancy	6 (6.9%)	84 (6.9%)	7 (7.1%)	100 (7.1%)	20 (11.7%)	250 (11.1%)
CLINICAL MEASURES						
Systolic BP (mm Hg)	166 [154, 180]	160 [151, 170]	167 [155, 180]	160 [151, 171]	160 [145, 176]	155 [145, 168]
Diastolic BP (mm Hg)	100 [93, 108]	100 [94, 105]	100 [93, 108]	100 [94, 105]	96 [88, 105]	96 [89, 103]
Uric acid	434 [367, 495]	376 [321, 429]	424 [347, 488]	377 [322, 430]	404 [327, 466]	369 [310, 427]
Chest pain/ dyspnoea**	20 (23.0%)	62 (5.1%)	22 (22.2%)	73 (5.2%)	25 (14.6%)	91 (4.0%)
Lowest platelet count ($\times 10^9$ per L)**	121 [56, 179]	177 [139, 219]	121 [67, 178]	179 [139, 221]	143 [90, 206]	190 [150, 235]
Highest AST (U/L)**	57 [33, 314]	33 [26, 46]	54 [31, 251]	32 [24, 44]	36 [23, 106]	28 [21, 41]
Creatinine	70 [61, 105]	63 [55, 73]	74 [61, 103]	64 [55, 74]	75 [30, 93]	64 [55, 75]

Characteristics	Temporal		Primary external		Broader	
	Outcome (87 women)	No outcome (1223 women)	Outcome (99 women)	No outcome (1405 women)	Outcome (171 women)	No outcome (2258 women)
INTERVENTIONS DURING ADMISSION						
Corticosteroids	27 (31.0%)	293 (24.0%)	30 (30.3%)	377 (26.8%)	59 (34.5%)	587 (26.0%)
Antihypertensive therapy	74 (85.1%)	822 (67.2%)	84 (84.9%)	974 (69.3%)	139 (81.3%)	1557 (69.0%)
$MgSO_4$	54 (62.1)	339 (27.7%)	61 (61.6%)	412 (29.3%)	110 (64.3%)	858 (38.0%)
PREGNANCY OUTCOMES						
Admission-To-Delivery Interval (Days)	1 [1, 1]	1 [1, 2]	1 [1, 1]	1 [1, 3]	1 [1, 1]	1 [1, 3]
Admission-To-Delivery Interval, <34^{+0} Weeks (Days)	1 [1, 2]	4 [1, 10]	1 [1, 2]	5 [2, 10]	2 [1, 3]	4 [2, 10]
Gestational age at delivery (wk)	36.4 [33.8, 38.6]	37.9 [36.1, 39.1]	36.4 [33.8, 38.6]	37.7 [35.9, 39.0]	35.9 [32.7, 38.1]	37.0 [34, 38.6]
C/S	60 (70%)	707 (58%)	70 (70.7%)	832 (59.2%)	126 (73.7%)	1360 (60.2%)
Birth weight	2465 [1752, 3088]	2900 [2295, 3388]	2410 [1818, 3050]	2855 [2222, 3350]	2277 [1475, 2959]	2680 [1840, 3240]
Neonatal death	0	10 (0.8%)	0	11 (0.8%)	1 (0.6%)	24 (1.1%)
Stillbirth	1 (1.2%)	6 (0.5%)	1 (1.0%)	6 (4.3%)	3 (1.8%)	22 (1.0%)

Table 4-5 Data completeness in the validation datasets

Variables	Temporal (N=1310)			Primary (N=1504)			Broader (N=2429)
	BCW (1310)	FINNPEC (124)	PELICAN (70)		PETRA (644)	Oxford (281)	
GA onset	0	0	0	0	0	2 (0.7%)	2 (0.1%)
Chest pain or dyspnoea	0	0	0	0	631 (98.0%)	273 (97.2%)	904 (37.2%)
SpO_2	323 (24.7%)	124 (100%)	12 (17.1%)	459 (30.5%)	290 (45.0%)	281 (100%)	1030 (42.4%)
Platelets	12 (0.9%)	1 (0.8%)	1 (1.4%)	14 (0.9%)	61 (9.5%)	6 (2.1%)	81 (3.3%)
AST or ALT	27 (2.1%)	6 (4.8%)	2 (2.9%)	35 (2.3%)	69 (10.7%)	6 (2.1%)	110 (4.5%)
Creatinine	28 (2.1%)	73 (58.9%)	1 (1.4%)	102 (6.8%)	72 (11.2%)	4 (1.4%)	178 (7.3%)

Table 4-6 Complete vs Incomplete for Primary external Dataset

Characteristics	Complete cohort (1394 women)	Incomplete cohort (110 women)	*P value*
DEMOGRAPHICS & PREGNANCY CHARATERISTICS			
Maternal age at EDD (yr)	34 [30, 38]	32.0 [29.0, 36.0]	0.0049
Parity ≥1	442 (31.7%)	23 (20.9%)	0.0243
Gestational age at eligibility (wk)**	37.4 [35.3, 39]	36.4 [32.9, 38.1]	0.0001
Multiple pregnancy	135 (9.7%)	6 (5.5%)	0.1952
Smoking in this pregnancy	98 (7.0%)	9 (8.2%)	0.7951
CLINICAL MEASURES			
Systolic BP (mm Hg)	160 [151, 171]	163 [150, 178]	0.3839
Diastolic BP (mm Hg)	100 [94, 105]	101 [93, 109]	0.1852
Uric acid	379 [323, 436]	364 [329, 421]	0.5603
Chest pain/dyspnoea**	84 (6.0%)	11 (10.0%)	0.1482
Lowest platelet count ($\times 10^9$ per L)**	174 [135, 217]	187 [153, 234]	0.0112
Highest AST (U/L)**	33 [25, 47]	20 [13, 29]	<0.0001
Creatinine	64 [56, 76]	54 [52, 61]	0.0495
INTERVENTIONS DURING ADMISSION			
Corticosteroids	368 (26.4%)	39 (35.5%)	0.0516
Antihypertensive therapy	975 (69.9%)	83 (75.5%)	0.2669
$MgSO_4$	421 (30.2%)	52 (47.3%)	0.0003
PREGNANCY OUTCOMES			
Admission-To-Delivery Interval (Days)	1 [1, 2]	3 [1, 7]	<0.0001
Gestational age at delivery (wk)	37.7 [35.9, 39.0]	37.0 [34.1, 38.3]	0.0017
C/S	832 (59.7%)	70 (63.6%)	0.4756
Birth weight	2852 [2210, 3345]	2580 [1810, 3045]	<0.0001
Neonatal death	11 (0.8%)	0 (0%)	1
Stillbirth	6 (0.4%)	1 (0.9%)	0.4130
MATERNAL OUTCOME (N women)			
Within 2 days	90 (6.5%)	9 (8.2%)	0.6150
Within 7 days	121 (8.7%)	31 (28.2%)	<0.0001

Characteristics	Complete cohort (1394 women)	Incomplete cohort (110 women)	*P value*
At anytime	145 (10.4%)	45 (40.9%)	<0.0001

Table 4-7 Risk stratification table for varying predicted probability cut-off values within 48h in the Primary External Validation dataset

Prediction score range	Total N women in range (%) (N=1504)	N women with outcome (%) (N=99)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	*TPR (%) [95% CI]	FPR (%) [95% CI]
<1.0%	459 (30.5%)	8 (1.7%)	0·25 [0·12-0.52]	-	-	-	-
1·0-2·4%	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%	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%	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%	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%	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. *TPR, True positive rate (or Sensitivity), FPR, false positive rate (1-Specificity)

Table 4-8 Risk stratification table at varying predicted probability cut-off values within 48h in the Temporal Validation 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]	TPR (%) [95% CI]	FPR (%) [95% CI]
<1.0%	395 (30.2%)	6 (1.5%)	0·22 [0·09-0·50]	-	-	-	-
1·0-2·4%	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%	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%	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%	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%	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]

Table 4-9 Risk stratification table at varying predicted probability cut-off values within 48h in the Broad Validation dataset

Prediction score range	Total N women in range (%) (N=2,429)	N women with outcome (%) (N=171)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	TPR (%) [95% CI]	FPR (%) [95% CI]
<1%	604 (24.9%)	14 (2.3%)	0·31 [0·18-0·55]	-	-	-	-
1·0-2·4%	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%	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%	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%	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%	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]

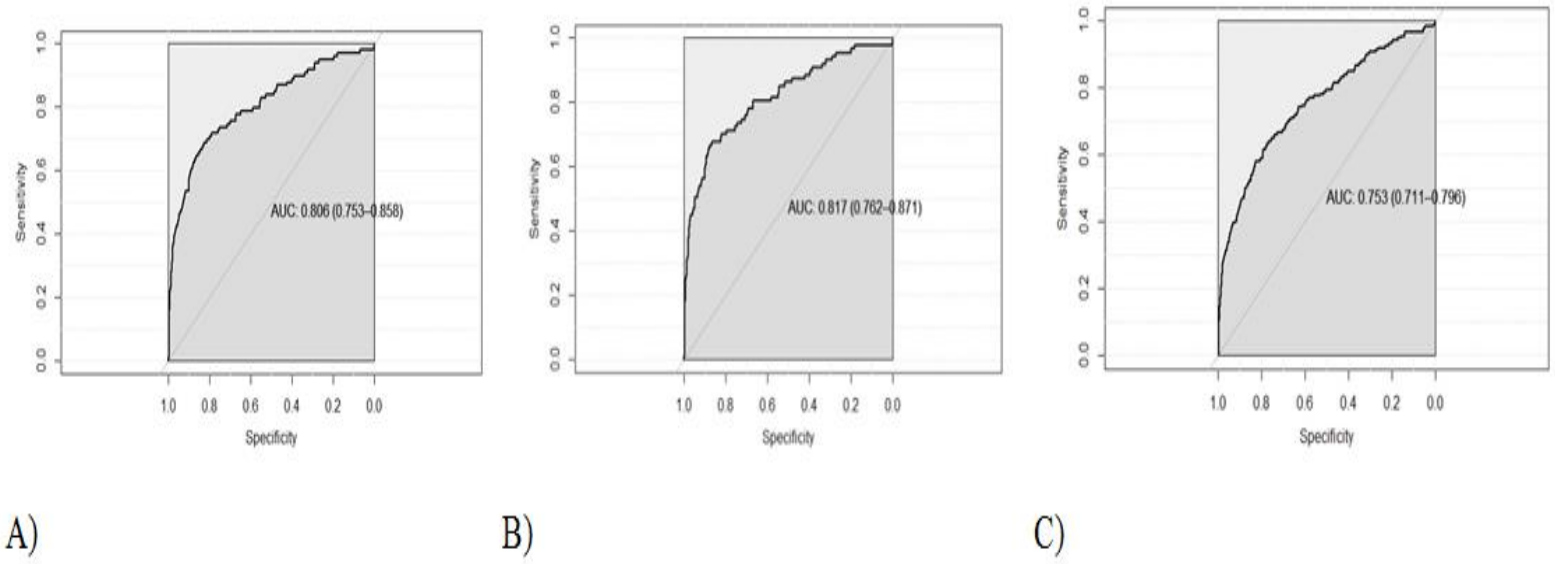


Figure 4-1 Discriminatory performances of the fullPIERS model in A) Primary external, B) Temporal dataset and C) Broader external datasets

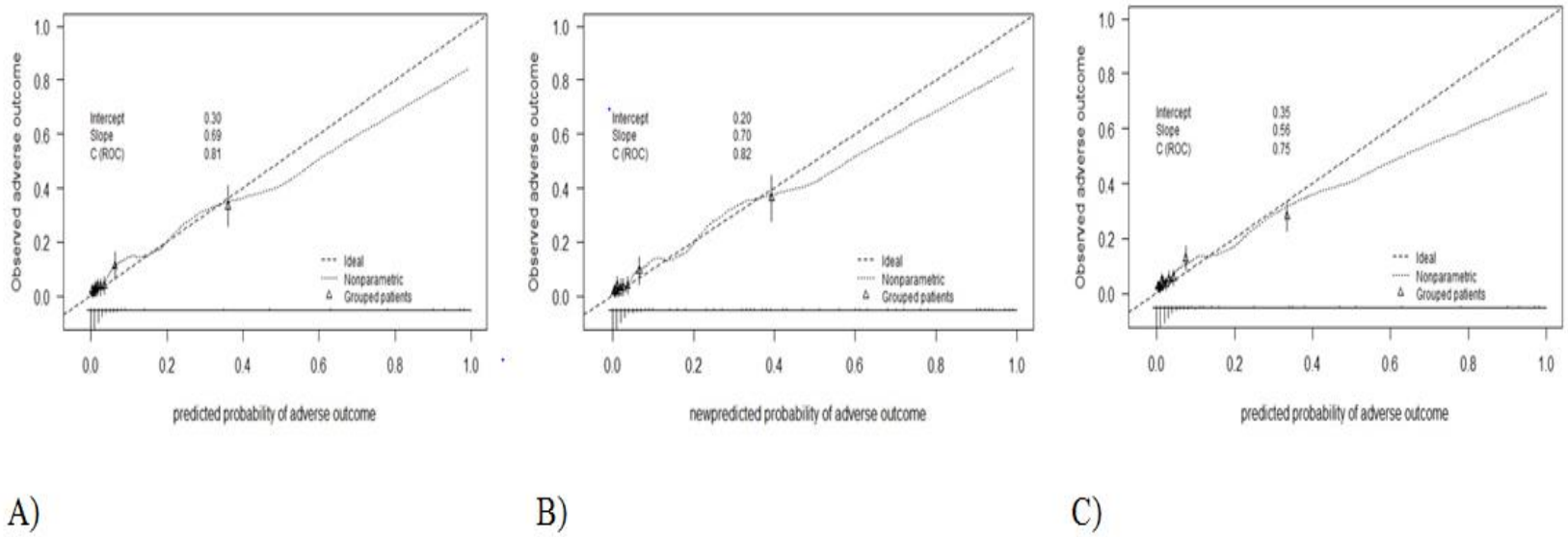


Figure 4-2 Calibration graphs of the fullPIERS model in the (A) Primary external (B) Temporal and (C) Broader validation datasets

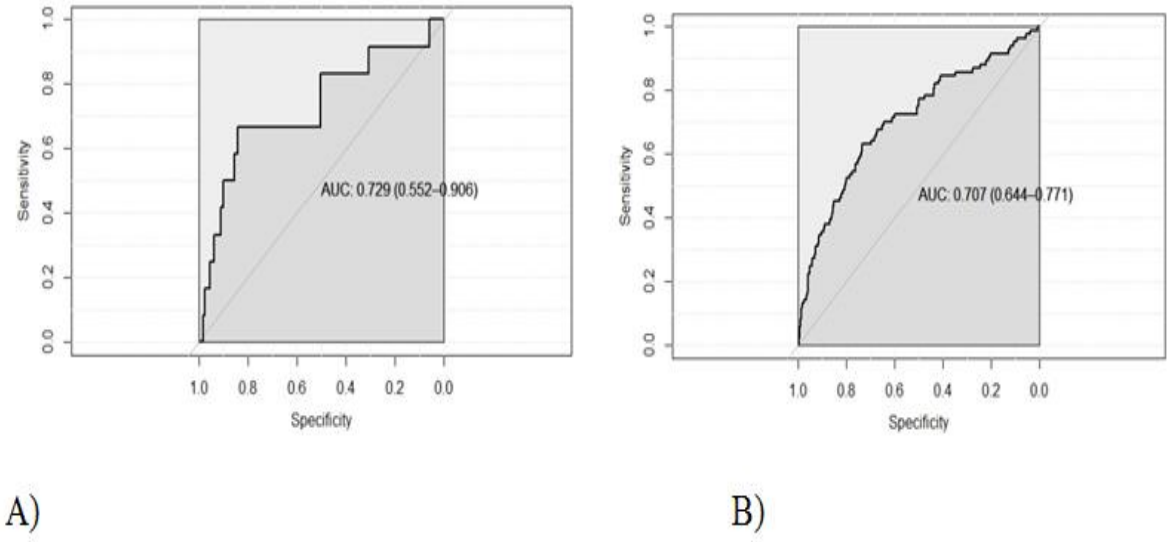


Figure 4-3 AUROCs of the fullPIERS model without the BCW cohort in A) Primary and B) Broader external datasets

Chapter 5: Domain (Broader Spectrum) assessment of the fullPIERS model

5.1 Background

Domain validation assesses the generalizability and potential use of the model across a broader spectrum of patients.[107,108] The different domains that can be assessed include a sub-group of patients with the disease, different definitions of disease and differences in predictor and outcome measurements.[107] Usually, a higher case-mix is expected because of these variations. In **section 5.2** of this chapter, the model is assessed in women admitted with early-onset pre-eclampsia in high-income countries and **in section 5.3**, in women with any HDPs admitted in in low-and-middle-income countries (LMICs).

5.2 Assessment of the fullPIERS model in women with early-onset pre-eclampsia

As discussed in **Chapter 1**, studies have proposed that early-onset pre-eclampsia (i.e. before 34 weeks of gestation) is related to shallow invasion of the maternal spiral arteries by the trophoblasts resulting in impaired remodeling of the arteries (placental pre-eclampsia).[13,14] Although late-onset pre-eclampsia is more common, early-onset pre-eclampsia is associated with more severe outcomes such as fetal growth restriction (FGR). [13,14] The management of early-onset pre-eclampsia is complicated since delivery remains the intervention that initiates the resolution of the clinical syndrome of pre-eclampsia and could result in very early preterm birth with the concomitant severe consequences of prematurity. Therefore, delaying delivery where possible would be preferable although the length of time for expectant management is unclear since the mother is also at increased risk of complications.

The ability to predict the risk of maternal complications for women admitted with early-onset pre-eclampsia would be highly beneficial to guide their management in all healthcare settings.[19]

Although the majority of the cohort used for the model development was from women with late-onset pre-eclampsia,[24] 31.4% of the women included in the study had early-onset pre-eclampsia. The objective of this section was to assess and confirm the validity of the fullPIERS model for early-onset pre-eclampsia, using a sufficiently powered, broad cohort of women admitted with early-onset preeclampsia in HICs.

5.2.1 Methods

5.2.1.1 Datasets

Data used for this model assessment study were derived from three pre-existing cohorts of women admitted with early-onset pre-eclampsia in HICs. These were the BCW cohort (ii) the Dutch-PETRA[90] and (iii) the PREP[91] cohorts, described in **Chapter 3**.

For this study, the BCW cohort was restricted to the women admitted with pre-eclampsia before 34 weeks of gestation. These cohorts were merged into a combined dataset of women admitted with early-onset pre-eclampsia.

5.2.1.2 Definition of pre-eclampsia and adverse outcomes

Pre-eclampsia was defined as in the fullPIERS development study.[15] However, the PETRA cohort included only women with severe pre-eclampsia (diastolic blood pressure (DBP) > 110 mm Hg). In addition, women having gestational hypertension (DBP ≥90 mm Hg in the absence of proteinuria) with FGR (estimated fetal weight <10th centile) were also included in the PETRA cohort.

The primary outcome used in our study was the same as in the model development study listed in the **Appendix B**.

5.2.1.3 Statistical analyses

Using the worst measured predictor variables within 48 hours of admission, the fullPIERS model equation[24] was applied to the combined dataset to calculate the predicted probabilities of experiencing an adverse outcome for each woman. The calculated probabilities were then used to assess the model performance for predicting adverse maternal outcomes within 48 h based on discrimination, calibration, and stratification and classification accuracy, using the same approach for interpreting results as in **Chapter 3**.

5.2.2 Results

5.2.2.1 Cohort description

The BCW, PETRA and PREP cohorts included 218, 216, and 954 women respectively, making a total of 1,388 women admitted with pre-eclampsia before 34 weeks of gestation in our analytical dataset. The women in the BCW cohort appeared to be older and have a higher rate of chest pain or dyspnea and more interventions during pregnancy (higher administration of corticosteroids, antihypertensive medication and magnesium sulphate [$MgSO_4$]) **(Table 6-1)**. The PETRA cohort had the highest rate of HELLP syndrome, and higher rates of stillbirth and neonatal death. The PREP cohort had the lowest rate of HELLP syndrome, higher multiparity and had lower interventions during pregnancy.

The PETRA cohort also had the highest rate of adverse maternal outcomes within 48 h of admission (14.8%) while the PREP cohort had the lowest rate (4.8%) (**Table 5-1**). In total, the rate of adverse outcomes in the combined dataset was 7.3% (N=101), which was slightly higher than in the fullPIERS cohort of 5%.

5.2.2.2 Model performance

The women in the PETRA cohort appeared to have a higher calculated fullPIERS probability (**Table 5-2**). The AUROC was also highest in the PETRA cohort (AUROC of 0. 97 (95% CI: 0.94-0.99)). The model showed a good discrimination in the combined data with AUROC of 0. 80 (95% CI: 0.75-0.86)) (**Figure 5-1**), although the calibration was poor with a slope of 0.68 (**Figure 5-2**).

The risk stratification capacity in the early-onset pre-eclampsia cohort remained comparable with to the model development study.[24] The fullPIERS model stratified the majority of the women (63.6%) into the low risk groups (predicted probability of < 2.5%) and 4.4% into the highest risk group (predicted probability of ≥ 30%) (**Table 5-3**). Conversely, only about 3% of women in the low risk group of <2.5% had an adverse outcome while approximately 55% of the women in the highest risk group experienced an adverse outcome. At the highest predicted probability group of ≥ 30%, the model had a LR+ of 23.4 (95% CI 14.83-36.79), showing strong evidence to rule in adverse outcome; the PPV and NPV were 96% and 65% respectively. There was no predicted range showing strong evidence for ruling out adverse outcomes.

5.2.3 Discussion

5.2.3.1 Main findings

The model maintained a good discriminatory and stratification performance in women admitted with early onset pre-eclampsia although there was a marginal decrease in the AUROC compared to the model performance in development (AUROC of 0. 80 (95% CI: 0.75-0.86) in early-onset pre-eclampsia vs 0.88 (95% CI: 0.84–0.92) on development). The calibration performance of the model decreased in our cohort from an ideal slope of 1 to 0.68.

The case-mix differences between our cohort and the original fullPIERS cohort may have attenuated the model's performance, particularly the calibration performance.[103,104] The most obvious case-mix difference was the selective inclusion of only women with early-onset pre-eclampsia compared with the fullPIERS cohort which had a higher proportion of women with late onset pre-eclampsia. In addition, earlier onset of pre-eclampsia (i.e., the gestational age of onset) were associated with more adverse outcomes as shown by the overall higher rate of outcomes in this dataset compared to the fullPERS cohort. Therefore, it is possible that the predictor effect of gestational age in the fullPIERS model may have been different in our cohort compared to the fullPIERS cohort. Differences in predictor effect can influence a model's performance, especially the calibration accuracy.[21,104] Other contributors to case-mix differences include the addition only of women admitted with severe pre-eclampsia as in the PETRA cohort compared with all women with pre-eclampsia in the model development, as well as the addition of women admitted into both secondary and tertiary units in the PREP cohort compared with those only admitted to tertiary units in the model development. Another possible reason for the overall model performance reduction is the lack of spread between low- and high-risk women in the combined data (i.e., less heterogeneity among the women in the combined cohort).[21,103,104] Despite these known differences, our primary goal was to assess how well the model would perform in this subset of pre-eclampsia in order to determine whether it would be useful for this population.

The AUROCs in all the individual datasets were good (≥ 0.70), although the discriminatory performance appeared to be higher in the PETRA dataset, which was even better than the original model performance. We suspect that the inclusion of a wider definition of women (i.e., women with gestational hypertension with FGR, women with severe-pre-eclampsia

and women with HELLP syndrome) may have resulted in more heterogeneity and more severity (also observed by the increased rate of adverse fetal outcomes) within this cohort, leading to the observed higher discrimination performance. In addition, this cohort had the highest proportion of HELLP syndrome and so it is possible that the effect of platelets in this group may have been higher compared to the other groups of women in the combined cohort.

5.2.3.2 Strengths and weaknesses

An important strength in our study is the combination of cohorts from different centres which added to the robustness and generalisability of our findings. We used a dataset with adequate sample,[102] enabling us to detect any true changes in the model performance. Since we were interested in assessing the model in a general population of early-onset pre-eclampsia, we believe that the combination of these cohorts resulted in a broader cohort of cases that could present clinically.

A possible limitation in our study is that we were not able to exclude the women with only gestational hypertension and FGR from the PETRA data, due to lack of availability of information, in order to test the model performance in the women with only early-onset pre-eclampsia. This may have provided information to test the proposed reasons stated above for heterogeneity in the data.

5.2.3.3 Comparison to existing literature

A prediction model study (PREP model)[91] on the prognosis of women with early onset pre-eclampsia reported an AUROC of 0.84 (95% CI, 0.81–0.87) upon development. However, the majority of the adverse outcomes predicted in the study were preterm deliveries (61.3%), which could be due to either fetal or maternal indications or both . This study by Thangaratinam *et al*[91] did not report any sensitivity analysis for the performance of the model in predicting other

adverse maternal outcome while excluding preterm delivery. Therefore, it is possible that this model may not be useful for the prediction of maternal complications. Another concern with the use of the PREP model is the inclusion of over 14 variables, making it cumbersome compared with the six-variable fullPIERS model. Model development studies have encouraged the use of a more parsimonious model as this reduces the chances of overfitting and enhances clinical utility.[21,109] Finally, the model in the study included treatment variables such as antihypertensive and $MgSO_4$; the administration and timing of these treatments may vary based on the clinician's training and experience. Therefore, we propose that the fullPIERS model might be a better for identifying women at highest risk regardless of treatment.

5.2.4 Conclusion

The fullPIERS model was able to discriminate well between women who experienced an adverse outcome and for early-onset pre-eclampsia. Our findings could guide decision-making especially the timing of delivery and planning of transfer to units for required care and administration of corticosteroid and $MgSO_4$ therapy. Thus, we believe that the fullPIERS model could aid in averting severe maternal complications as well as the reduction of unnecessary preterm delivery.

5.3 Assessment of the fullPIERS model in women with HDPs in LMICs

As stated in **Chapter 1**, the burden of the HDPs is higher in the low- and middle-income countries (LMICs), where greater than 90% of HDP-related deaths occur.[13,14] Therefore the model's potential for use in a LMIC setting, where the majority of HDP-related morbidity and mortality occur, was assessed.

5.3.1 Methods

5.3.1.1 Dataset

The miniPIERS cohort,[62] described **in Chapter 3**, was used for this model assessment study. In summary, the women in this cohort were admitted with a HDP and collected prospectively in LMICs.

5.3.1.2 Definition of HDPs and adverse outcomes

Pre-eclampsia was defined as in the fullPIERS cohort;[24] gestational hypertension was defined as blood pressure (BP) ≥140/90 mmHg (at least one component, twice, ≥4 hours apart, $\geq 20^{+0}$ weeks) without significant proteinuria, and chronic hypertension as BP ≥140/90 mmHg (at least one component, twice, ≥4 hours apart, $<20^{+0}$ weeks' gestation). Adverse maternal outcomes were defined as in the fullPIERS study **(Appendix A)**.

5.3.1.3 Statistical analyses

The same statistical methods for handling missing data, calculating the predicted probabilities for each woman, and evaluation of model performance and interpretation of performance, as described in **Chapter 3**, were used. The model intercept was updated to account for the baseline differences in the rate of adverse outcomes before model performance evaluation.[24]

5.3.2 Results

5.3.2.1 Cohort description

Of the 2,081 women in the miniPIERS cohort, 261 (12.5%) developed an adverse maternal outcome(s) within 48 hours of hospital admission with a HDP. Seven hundred and fifty-seven (36.4%) women had information for all variables in the fullPIERS model and these women were used for this validation study (complete case analysis).

Of the 757 complete cases, 109 (14.4%) women had an adverse maternal outcome(s) within 48 hours of hospital admission. The most common adverse outcomes encountered were blood transfusion (N=52), eclampsia (N=14), and pulmonary oedema (N=18). Other notable outcomes are listed in Table 5-4. There was no case of maternal death recorded in the validation dataset. Women in the miniPIERS validation cohort vs. the fullPIERS development cohort were different with regards to demographics and pregnancy characteristics (i.e., slightly younger, more often parous, and less likely to be a smoker or have a multiple pregnancy), clinical measures (i.e., lower dipstick proteinuria, lower platelet count, and lower creatinine), interventions (i.e., more likely to receive antenatal corticosteroids, antihypertensive therapy, and $MgSO_4$), and outcomes (i.e., shorter admission to delivery interval, higher infant birth weight but a higher infant mortality before hospital discharge) (**Table 5-5**).

Within the miniPIERS validation dataset, women who had adverse outcomes (vs. those who did not) were slightly younger, more often nulliparous, and had hypertensive disorders of greater severity, including higher BP, more frequent antihypertensive therapy and $MgSO_4$, early gestational age at delivery, and lower infant birth weight compared to women without an adverse outcome (**Table 5-6**).

5.3.2.2 Data completeness and imputation analysis

Seven hundred and fifty-seven (36.4%) women (568 (75%) with pre-eclampsia and 189 with other HDPs) in the miniPIERS dataset had complete fullPIERS variables. All women in the miniPIERS cohort had data for the gestational age at eligibility and chest pain/dyspnoea; missing SpO_2 values (1423, 68.3%) were substituted with 97% , as had occurred in the fullPIERS model development[21] and multiple imputations were carried out for missing platelet count (1297,

62·3%), serum creatinine (1282, 61.6%), and AST (923, 44.4%). Imputation of missing values did not appear to alter the model performance significantly.

5.3.2.3 Model performance

The fullPIERS model predicted an adverse maternal outcome within 48 hours of eligibility in the miniPIERS validation cohort with good discriminative performance as indicated by an AUROC of 0.77 (95% CI: 0.72-0.82).[24] There was no significant change in the model performance using only cases with pre-eclampsia.

Figure 5-4 shows the calibration plot of the fullPIERS model when applied to the miniPIERS validation cohort. The calibration performance of the model was poor with a slope of 0·67 and intercept of -0.53 showing underestimation of risk at the lower risk ranges and overestimation of risks at the high-risk ranges. **Table 5-7** presents tabular information about stratification and classification accuracy. In the fullPIERS development cohort, more women (35%) fell into the predicted risk category of <1.0% than any other category, whereas in the miniPIERS complete-case validation cohort, the 5.0-9.9% range was the most common (with 23.5% of women). The majority of women who experienced an adverse outcome in both cohorts were in the predicted risk category of ≥ 0.30 (i.e., 59% for fullPIERS and 50% for miniPIERS). Thus, the model classified a greater proportion of women without outcomes into the middle group, indicating lower stratification accuracy for the low-risk groups; classification accuracy remained good for the high-risk group in the validation cohort.

Using the highest predicted probability cut-off of 0.30, the category into which most women with adverse outcomes fell, the likelihood ratio was moderate at 5.9 (95% CI: 4·2-8.4) with a PPV of 50% (95% CI: 0.40-0.60). Overall, the negative predictive values remained high (> 90%) across all the risk due to the low prevalence of outcomes (**Table 5-7**).

5.3.3 Discussion

5.3.3.1 Main findings

The model had good discriminative ability with an AUROC of 0.77 (95% CI: 0.72-0.82) within 48 hours of admission, using the miniPIERS cohort of women in low-resourced settings for the prediction of adverse maternal outcomes related to the HDP, but this was significantly lower than its original performance in the development cohort (AUROC 0.88, 95% CI: 0.84-0.92). Despite updating the model intercept to account for the baseline differences in adverse outcomes between the development and validation cohorts, the fullPIERS model had a poor fit in the miniPIERS dataset reflected by the poor calibration performance. However, the fullPIERS model performed moderately as a 'rule-in' test in the highest probability risk group with likelihood ratio of 5.9 (95% CI: 4.23 – 8.35).

A possible reason for the decrease in the performance of the fullPIERS model in our study was the heterogeneity between the development cohort and our validation cohort.[21,103] Differences between the development cohort and our validation cohort existed in the inclusion criteria, outcome prevalence, data collection settings (high-resourced vs low-resourced countries), and predictor distribution such as AST and platelet count. Such low- and middle-income settings as our validation cohort settings are more likely to have more co-morbidity, lower socioeconomic status, less availability of resources and differences in disease management compared with high-income settings (reflected by more co-interventions and the shorter admission-to delivery interval in the validation cohort shown in **Table 5-5**). Such factors may result in case-mix differences, and may also alter the effect of the predictors on the outcome.[103,104] Therefore, the extreme predictions observed in the calibration slope may have

been as a result of differences in the predictor effects in the validation and development cohort.[103,104] These factors may have resulted in the reduced performance of the model.

5.3.3.2 Strengths and weaknesses

A strength of this study is the assessment of the fullPIERS model in a broader population (i.e., in a low-resourced setting with any HDP) using a sufficiently powered sample size.[102] While internal and external validation using a similar patient cohort are important, validating a model in a different broader spectrum is needed to evaluate the generalizability of the prediction model in other settings with a more diverse group of patients. This external validation study using data from LMICs is particularly useful as most of the global burden of mortality and morbidity from the HDPs is borne by low-resourced settings.

The observed LR (5.9) at the highest classification group suggests that the fullPIERS model can be used as a moderate 'rule-in' tool for adverse outcomes from pre-eclampsia and other HDPs in low-resourced settings. For clinical practice in these settings, the recommended predicted probability of 0.3 can also be used as the optimal cut-off point to guide decisions around the need for immediate interventions. Half of the women with an adverse outcome fell in this risk category while the model still maintained a good likelihood ratio at a low false positive rate (7.6%). This has the added advantage of focusing limited resources on those who most need assistance in LMICs.

The major limitation of this study is the high proportion of missing values since the miniPIERS data were not originally collected explicitly for the purpose of this study. Using only complete-case analysis can lead to biased estimates of the predictions if the validation subset is not truly representative of the population at risk. Imputation of missing values did not show any significant change in the model performance. Therefore, it was less likely that selection bias

contributed significantly to the decreased performance of the fullPIERS model in the complete case analysis compared with the development performance.[92] Even when missing values were excluded, the complete-case analysis had sufficient power (109 outcomes) to externally validate the model.[102]

Of note, most of the variables were missing most likely because laboratory measurements for pre-eclampsia and the other HDPs are usually ordered based on the severity of other clinical measurements.[11] This clinical management practice reflects the scarcity of resources in LMIC public hospitals and should draw attention to the need for a lower cost of point-of-care laboratory measurement techniques for these important laboratory measures. In this validation cohort, there were worse clinical measures and pregnancy outcomes observed in the women with complete laboratory data compared to those with missing laboratory results (**Table 5-8**). This suggests that clinicians in these settings are able to identify some of the higher risk women based on clinical assessment alone but that there remains a delay in timely intervention, so women continue to experience poor outcomes; it could also be that the women were sicker by the time they arrived to the healthcare centre.[11] Reducing the delay between assessments of laboratory measures and intervening when indicated should improve these women's outcomes.

5.3.3.3 Comparison with other findings

A study by Agrawal and Maitra[98] assessed the validity of the fullPIERS model in a low-income setting using a cohort of 323 women admitted with pre-eclampsia. The study cohort had a higher rate of adverse outcomes (18.5%) compared to the original miniPIERS cohort (14.4%), included only pre-eclampsia versus any HDPs in the miniPIERS and used less expectant management in the their study compared to the miniPERS and fullPIERS cohorts. However, they also reported a

similar decrease in the model performance with AUROC of 0.71 (95% CI: 0.67 to 0.74) in their cohort.

The miniPIERS model which was developed using this cohort (miniPIERS cohort), using demographics, signs and symptoms of HDPs reported an AUROC of 0.71 (95% CI: 0.66–0.77) upon external validation. Although the LR was not reported for the external validation of the miniPIERS model, the performance in the original model was slightly lower than the fullPIERS model in the highest risk group. Therefore, the fullPIERS model may be used when laboratory tests are present to provide more accurate predictions in LMICs.

5.3.3.4 Interpretation

The fullPIERS model showed moderate utility for the prediction of adverse maternal outcomes in women with HDPs in our validation cohort collected in low-resourced setting hospitals. The stratification accuracy and discriminative ability of the fullPIERS model within the highest risk group makes it a valuable tool to aid clinicians in the identification of women at highest risk of adverse outcomes and allow for timely delivery of appropriate interventions such as transfer to a higher level of care for delivery, and administration of antenatal corticosteroids. To determine applicability of the model in other well-resourced settings, future validation studies using more similar cohorts to that in which the model was developed are still needed.

5.4 Summary

The fullPIERS model had good discriminatory and stratification performance in broader spectrum validation including women with early-onset pre-eclampsia, other HDPs and admitted into different units in other settings. However, calibration performances of the model in these settings were significantly decreased, suggesting the need for more extensive updating of the model to improve its performance in these settings.

Table 5-1 Maternal characteristics for the datasets with women GA <34 (BC Women, Dutch PETRA, PREP)

Characteristics	BCW (218 women)	Dutch PETRA (216 women)	PREP (954 women)
DEMOGRAPHICS & PREGNANCY CHARACTERISTICS			
HELLP SYNDROME	27 (12.4%)	93 (43%)	10 (1.0%)
Maternal age at EDD (yr)	35 [30, 39]	30 [27, 34]	30 [26, 35]
Parity ≥1	84 (31.2%)	65 (30.1%)	403 (42.2%)
Gestational age at eligibility (wk)**	31.0 [28.4, 32.7]	30.0 [27.4, 31.4]	31.4 [28.7, 32.7]
Multiple pregnancy	40 (18.4%)	-	84 (8.8%)
Smoking in this pregnancy	24 (11.1%)	-	87 (9.1%)
CLINICAL MEASURES			
Systolic BP (mm Hg)	161 [150, 173]	160 [145, 170]	155 [145, 169]
Diastolic BP (mm Hg)	100 [94, 106]	105 [95, 110]	99 [92, 105]
Chest pain/dyspnoea**	27 (12.4%)	15 (6.9%)	60 (6.3%)
Lowest platelet count ($\times 10^9$ per L)**	189 [133, 235]	164 [89, 227]	222 [176, 273]
Highest AST/ALT (U/L)**	37 [27, 65]	32 [24, 46]	18 [13, 28]
Creatinine	64 [56, 78]	69 [58, 79]	59 [50, 69]
INTERVENTIONS DURING ADMISSION			
Corticosteroids	195 (89.5%)	153 (70.8%)	589 (61.7%)
Antihypertensive therapy	197 (90.4%)	171 (79.2%)	125 (13.1%)
$MgSO_4$	167 (76.6%)	89 (41.2%)	120 (12.6%)
PREGNANCY OUTCOMES			
Admission-To-Delivery Interval, (Days)	3 [1, 8]	8 [4, 15]	9 [3, 23]
Gestational age at delivery (wk)	32.0 [29.3, 33.6]	31.4 [28.3, 33.0]	33.1 [31.0, 34.9]
Birth weight	1340	1203	1625

Characteristics DEMOGRAPHICS & PREGNANCY CHARACTERISTICS	BCW (218 women)	Dutch PETRA (216 women)	PREP (954 women)
	[895, 1785]	[839, 1506]	[1260, 2165]
Stillbirth	6 (2.8%)	20 (9.3%)	16 (1.7%)
Neonatal death	3 (1.4%)	18 (8.3%)	23 (2.4%)

Table 5-2 PIERS prediction and outcomes in Early-Onset pre-eclampsia cohorts

Characteristics	BCW (218 women)	Dutch PETRA (216 women)	PREP (954 women)
MATERNAL OUTCOME (N women)			
Within 48h	23 (10.6%)	32 (14.8%)	46 (4.8%)
Within 7 days	36 (16.5%)	62 (28.7%)	81 (8.5%)
At anytime	46 (21.1%)	73 (33.8%)	103 (10.8%)
fullPIERS probability median (IQR)	0.0253 [0.0092, 0.0794]	0.0312 [0.0149, 0.1188]	0.0095 [0.0046, 0.0193]
fullPIERS probability mean (SD)	0.0138 (0.2104)	0.1387 (0.2390)	0.0237 (0.0664)
AUROC within 48 hrs (95% CI)	0.729 (0.595-0.863)	0. 970 (0.943-0.997)	0.730 (0.645-0.815)

Table 5-3 Risk stratification table at varying predicted probability cut-off values within 48h in the Early-Onset Pre-eclampsia

Prediction score range	**Total N women in range (%) (N=1388)**	**N women with outcome (%) (N=101)**	**LR [95% CI]**	**NPV (%) [95% CI]**	**PPV (%) [95% CI]**	***True positive rate (%) [95% CI]**	**False positive rate (%) [95% CI]**
<1.0%	594 (30.5%)	14 (1.7%)	-	-	-	-	-
1·0-2·4%	409 (33.1%)	17 (2.8%)	0.55 [0.36-0.86]	97.6 [96.0-98.7]	11.0 [8.9-13.4]	86.1 [77.5- 91.9]	54.9 [52.2-57.6]
2·5-4·9%	158 (19.1%)	8 (4.5%)	0.68 [0.34-1.34]	96.9 [95.6-97.9]	18.2 [14.5-22.5]	69.3 [59.2- 77.9]	24.5 [22.2-26.9]
5·0-9·9%	91 (7.8%)	6 (13.7%)	0.90 [0.40-2.01]	96.6 [95.4-97.6]	27.3 [21.7-33.7]	61.4 [55.1- 70.8]	12.8 [11.1-14.8]
10·0-29·9%	68 (5.1%)	12 (15.6%)	2.73 [1.51-4.92]	95.9 [94.7-96.9]	38.5 [30.2-47.4]	49.5 [39.5- 59.6]	6.2 [5.0-7.7]
≥30·0%	68 (4.4%)	44 (54.5%)	23.4 [14.83-36.79]	95.7 [94.4-96.7]	64.7 [52.1-75.6]	43.6 [33.8- 53.8]	1.9 [1.3-2.8]

One or More of Maternal Morbidity or Mortality within 48 h: Total *n* (%)	**Validation Cohort (*N* = 757) Outcomes *n* = 109 (14.4%)**
CENTRAL NERVOUS SYSTEM	
Eclampsia	14
Glasgow coma score <13	5
Stroke	1
Cortical blindness or retinal detachment	2
CARDIORESPIRATORY	
Positive inotropic support	1
Infusion of a 3rd parenteral antihypertensive	4
Myocardial ischaemia/infarction	3
SpO_2 <90%	5
Intubation (other than for cesarean section)	3
Pulmonary oedema	18
HAEMATOLOGICAL	
Transfusion of any blood product	52
Platelets $<50 \times 10^9$/l with no transfusion	5
HEPATIC	
Dysfunction	4
RENAL	
Acute renal insufficiency	3
Dialysis	1
PLACENTAL OUTCOMES	
Placental abruption	2

Table 5-4 Maternal adverse outcomes occurring in the complete-case miniPIERS cohort

Table 5-5 Maternal characteristics for the development (fullPIERS) and validation (complete-case miniPIERS) cohorts

Characteristics (N (%) or median [interquartile range])	**miniPIERS cohort (complete cases, validation) (757 women)**	**fullPIERS cohort (development) (2,023 women)**
DEMOGRAPHICS & PREGNANCY CHARATERISTICS		
Maternal age at EDD (yr)	28 [24, 33]	31 [27, 36]
Parity ≥1	406 (53·6%)	581 (28.7%)
Gestational age at eligibility (wk)**	36·6 [33.1, 38.1]	36 [33, 38.3]
Multiple pregnancy	18 (2.4%)	192 (9.5%)
Smoking in this pregnancy	48 (6.3%)	249 (12.3%)
CLINICAL MEASURES		
Systolic BP (mm Hg)	160 [150, 170]	160 [150, 176]
Diastolic BP (mm Hg)	100 [100, 110]	102 [97.8, 110]
Worst dipstick proteinuria	+2 [+1, +3]	+2 [+1, +4]
Chest pain/dyspnoea**	30 (4.0%)	90 (4.4%)
Lowest platelet count ($\times 10^9$ per L)**	187 [150, 231]	192 [150, 241.5]
Highest AST (U/L)**	30 [20, 35]	28 [21, 41]
Creatinine**	54 [10, 71]	67 [58, 77]
INTERVENTIONS DURING ADMISSION		
Corticosteroids	253 (33.4%)	550 (27.2%)
Antihypertensive therapy	704 (92.9%)	1381 (68.3%)
$MgSO_4$	376 (49.7%)	690 (34.1%)
PREGNANCY OUTCOMES		
Gestational age at delivery (wk)	37.1 [34.4, 38.6]	36.9 [34.1, 38.6]
Birth weight (g)	2500 [1896, 2433]	2141 [1441, 2807]
Infant death (before discharge)	26 (3.4%)	26 (1.3%)

Table 5-6 Demographics of the miniPIERS complete-case validation cohort according to adverse maternal outcomes

Characteristics (N (%) or median [interquartile range])	Women with an adverse outcome (109 women)	Women without an adverse outcomes (648 women)
DEMOGRAPHICS		
Maternal age at EDD (yr)	27 (±5.82)	29 (±6.46)
Parity ≥1	46 (42.2%)	360 (55.6%)
Gestational age at eligibility (wk)	36·5 [31.3, 37.9]	36.6 [33.4, 38.2]
Multiple pregnancy	1 (0.9%)	17 (2.6%)
Smoking in this pregnancy	4 (3.7%)	44 (6.8%)
CLINICAL MEASURES (WITHIN 24 HR OF ELIGIBILITY)		
Systolic BP (mm Hg)	170 [156, 190]	151 [145, 170]
Diastolic BP (mm Hg)	110 [107, 120]	100 [100, 110]
Worst dipstick proteinuria	+3 [+2, +3]	+2 [+0.5, +3]
INTERVENTIONS AT ANY TIME DURING ADMISSION		
Corticosteroids	35 (32.1%)	218 (33.6%)
Antihypertensive therapy	106 (97.2%)	598 (92.3%)
$MgSO_4$	83 (76.1%)	293 (45.2%)
PREGNANCY OUTCOMES		
Admission-to-delivery interval (d)	1 [1, 1]	2 [1, 5]
GA on delivery (wk)	36.6 [31.3, 38.1]	37.1 [34.7, 38.1]

Characteristics (N (%) or median [interquartile range])	**Women with an adverse outcome (109 women)**	**Women without an adverse outcomes (648 women)**
Birth weight (g)	2390 [1380, 2820]	2500 [1950, 3000]
Infant death before discharge	4 (3.7%)	22 (3.4)

Table 5-7 Risk stratification table at varying predicted probability cut-off values within 48h in the miniPIERS cohort

Prediction score range	Total in Range (757 women)	Adverse outcome (109 women)	LR [95% CI]	NPV (%) [95% CI]	PPV (%) [95% CI]	*True positive rate (%) [95% CI]	False positive rate (%) [95% CI]
<1.0%	30 (4.0%)	2 (6.7%)					
1.0-2.4%	107 (14.1%)	3 (2.8%)	0.17 [0.06-0·53]	93 [0·76-0·99]	15 [0.12-0.18]	98 [0.93- 0.99]	95.6 [0.94-0.97]
2.5-4.9%	140 (18.5%)	12 (8.6%)	0.56 [0.32-0.97]	96 [0.91-0·99]	17 [0.14-0.20]	95.4 [0.89-0.98]	79.6 [0.76- 0.83]
5.0-9.9%	178 (23.5%)	8 (4.5%)	0.28 [0.14-0.55]	94 [0.90-0.96]	19 [0.16-0.23]	84.4 [0.76-0.90]	59.9 [0.56-0.64]
10.0-29.9%	204 (27.0%)	35 (17.2%)	1.23 [0.91-1.67]	95 [0.92-0.96]	28 [0.23-0.33]	77.1 [0.68-0.84]	33.6 [0.30-0.37]
≥30.0%	98 (12.9%)	49 (50%)	5.9 [4.23-8·35]	91 [0.88-0.93]	50 [0.40-0.60]	45 [0.36-0.55]	7.6 [0.06-0.10]

Table 5-8 Demographics of the complete-case miniPIERS (n= 757) and the incomplete (n= 1324) subsets of the miniPIERS data

Characteristics	Complete-case subset (*n*= 757 women)	Incomplete subset (*n* = 1324 women)	*p*-Value*
DEMOGRAPHICS			
Maternal age at EDD (years) median [interquartile range]	28 [24, 33]	28 [24, 32]	0·5021
Parity ≥1 n (%)	406 (53.6%)	716 (54.1%)	0·8444
Gestational age at eligibility (wk) median [interquartile range]	36·6 [33.1, 38.1]	37.1 [33.7, 39.0]	<0·0001
Multiple pregnancy n (%)	18 (2.4%)	56 (4.2%)	0·0335
Smoking in this pregnancy n (%)	48 (6.3%)	49 (3.7%)	0·0083
CLINICAL MEASURES (WITHIN 24 H OF ELIGIBILITY)			
Systolic BP (mm Hg) median [interquartile range]	160 [150, 170]	160 [140, 171]	0·0124
Diastolic BP (mm Hg) median [interquartile range]	100 [100, 110]	100 [90, 114]	<0·0001
Worst dipstick proteinuria median [interquartile range]	+2 [+1, +3]	+1 [Trace, +3]	<0·0001
Chest pain/dyspnoea n (%)	30 (4·0%)	64 (4.8%)	0·4176

Characteristics	Complete-case subset (*n*= 757 women)	Incomplete subset (*n* = 1324 women)	*p*-Value*
INTERVENTIONS AT ANY TIME			
DURING ADMISSION			
Corticosteroid administration n (%)	253 (33.4%)	452 (34.1%)	0·7760
Antihypertensive medications administered n (%)	704 (92.9%)	1242 (93.8%)	0·5304
$MgSO_4$ administered n (%)	376 (49.7%)	572 (43.2%)	0·0050
PREGNANCY OUTCOMES			
GA on delivery (wk) median [interquartile range]	37·1 [34.4, 38.6]	37.6 [34.7, 39.3]	<0·0001
Birth weight (g) median [interquartile range]	2500 [1896, 2433]	2654 [1900, 3200]	0·0142
Infant death (before discharge) n (%)	26 (3.4%)	42 (3.2%)	0·8448

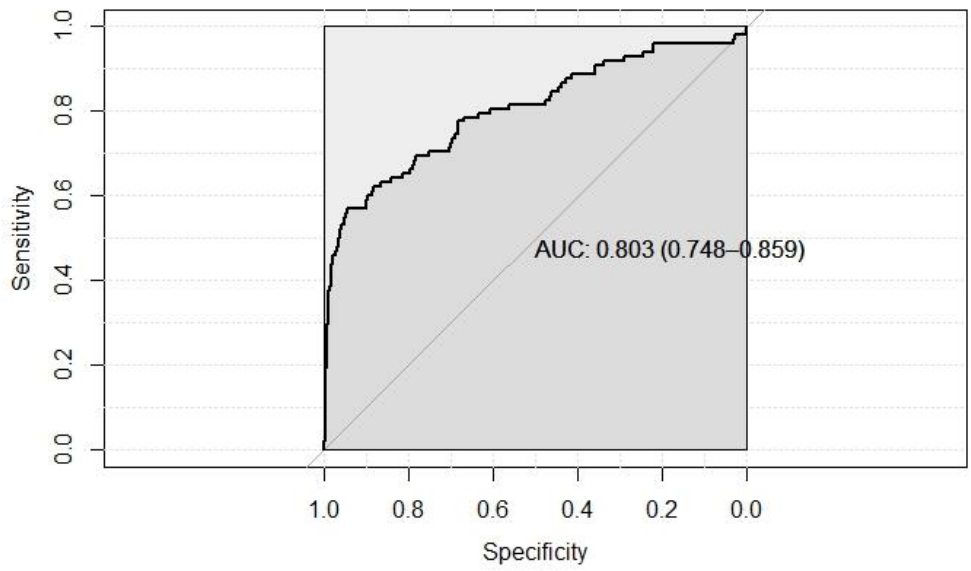


Figure 5-1 Discrimination performance of the fullPIERS model in early-onset pre-eclampsia within 48h of admission

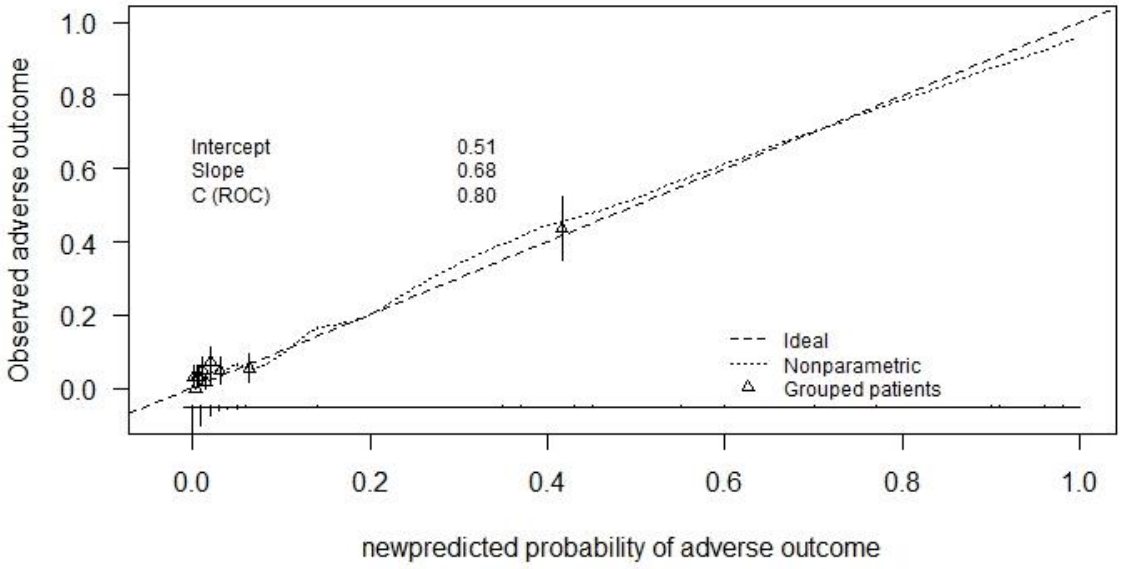


Figure 5-2 Calibration performance of the fullPIERS model in early-onset pre-eclampsia within 48h of admission

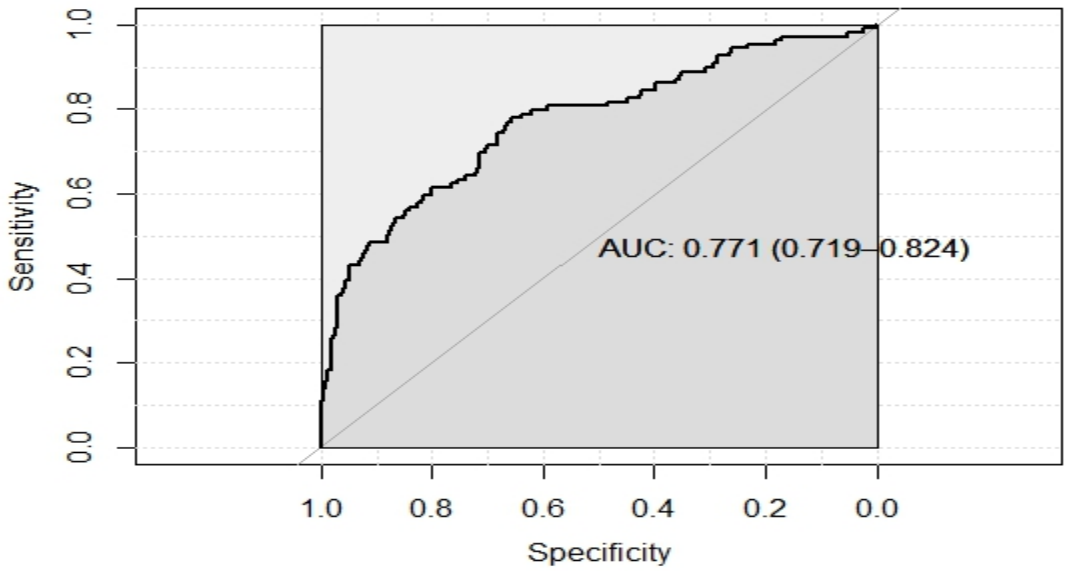


Figure 5-3 Discrimination performance of the fullPIERS model in the miniPIERS cohort within 48h of admission

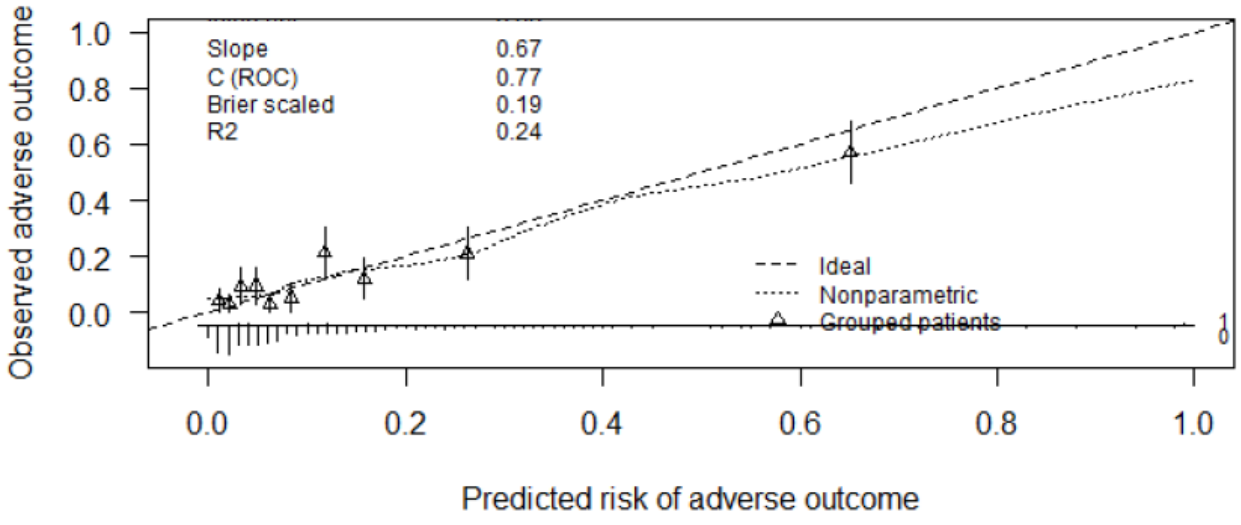


Figure 5-4 Calibration performance of the fullPIERS model in the miniPIERS cohort within 48h of admission

Chapter 6: Recalibration of the fullPIERS model and assessing the incremental value of the addition of PlGF to the model

6.1 Background

As shown in **Chapters 4** and **5**, the discriminatory performance of the model remained good upon validation assessment, especially for the temporal (AUROC 0.82 (95% CI: 0.76-0.87)) and primary (AUROC 0.81 (95% CI: 0.75-0.86)) validations. However, there was a non-significant decrease in the discriminatory performance and a decrease in the calibration performance especially in the external validation (intercept of 0.3 and slope of 0.7) compared with the original fullPIERS model on development and internal validation.[24] These trends of reduced performances suggest that there might be a need to update the model to improve its use and provide a more accurate prognosis to inform clinicians on the management of women with pre-eclampsia.[21,110] In addition to updating a model due to reduced performance, model updates are also generally required continuously due to changes in disease trends and practice over time, discovery of new techniques or biomarkers that might have prognostic ability, and for the expansion of model use.[21,107,111]

Updating methods range from simpler methods such as either recalibration of the intercept and/or slope or more complex methods such as re-estimation of some or all of the model variable coefficients and addition of new biomarkers.[21] The choice of updating usually depends on the extent to which the model performance improvement is required as well as the availability of new markers that might improve the performance.[107,111] While continuous model updating is important, the simpler methods are recommended to avoid overfitting of the model in the new datasets which might reduce transportability.[21] In **Section 2** of this Chapter, recalibration

(updating of the intercept) and logistic recalibration (updating of both the intercept and slope) are explored to improve the model performance.

Although little much has changed in terms of the management of pre-eclampsia and incidence of adverse outcomes since the development of the fullPIERS model in 2010, there has been some advancement in the discovery of biomarkers that could potentially aid in the risk identification of adverse outcomes. One of such biomarkers is placental growth factor (PlGF), which an angiogenic factor found in the maternal circulation.[70-73] Studies have shown than plasma concentration levels of PlGF are decreased in pregnancies complicated by pre-eclampsia compared to uncomplicated pregnancies.[67,68] This has led to further research on the potential use of PlGF as a diagnostic and prognostic marker for pre-eclampsia, especially for early onset pre-eclampsia.[70,72] As discussed in **Chapter 2.3**, some studies have shown that PlGF may have good potential as a prognostic test for adverse fetal outcomes associated with pre-eclampsia especially for preterm delivery and the delivery of small-for-gestational age (SGA) babies; [58,77,79-83] however, few studies have evaluated its prognostic ability for adverse maternal outcomes. The study by Ghosh *et al*[57] reported an increased risk of developing postpartum haemorrhage (PPH) in women with pre-eclampsia with low PlGF between 22 and 24 weeks of gestation. Some other studies reported high discriminatory performance or strong likelihood ratios (LRs) for combined adverse maternal and fetal outcomes in women with suspected pre-eclampsia.[61;80;84-85;87] Based on these reports, we examined whether or not the addition of PlGF to the fullPIERS model would provide better prognosis of maternal complications.

The benefits of adding PlGF to the fullPIERS model is unknown. In **Section 2** of this Chapter, the incremental value of PlGF (if any) to the fullPIERS model for prediction of adverse maternal outcomes is assessed.

6.2 Recalibration

6.2.1 Methods

6.2.1.1 Dataset

The three cohorts that were used for the primary external validation of the model in the previous chapter were used for the recalibration of the model intercept and slope: (i) BCW (ii) FINNPEC, and (iii) PELICAN.

6.2.1.2 Statistical analyses

Using the worst measured model predictor variables within 48 hours of admission, a linear predictor variable was calculated for all the women in the recalibration dataset with the fullPIERS model equation. To update the intercept (i.e. correcting the mean predicted outcome to equal the mean observed outcome), the linear predictor was used as an offset variable (variable with a fixed slope) in a logistic regression model, and a new intercept was generated.[21] The new intercept was used to calculate the updated predicted probabilities of experiencing an adverse outcome for each woman (new probabilities with updated intercept – recalibrated model A).
To update both the intercept and the overall slope, the linear predictor was used as the only co-variable in a logistic predictor regression. The recalibrated model equation (with a new intercept and coefficient) was then used to calculate the recalibrated predicted probabilities of experiencing an adverse outcome for each woman (probabilities with updated intercept and slope - recalibrated model B).

The updated model performances were re-assessed using discrimination, calibration, and stratification and classification accuracy, as described in **Chapter 4** and compared to the performance using the original fullPIERS model.[24]

6.2.2 Results

The updating of the model intercept (Recalibrated model A) resulted in a new intercept of 0.2023 while the Recalibrated model B (updated intercept and slope) resulted in an intercept of -0.5206 and an overall slope of 0.6963 **(Box 5-1)**.

Box 5-1:

Recalibrated linear predictor (A) = 0.2023 + original fullPIERS model predictor

Logistic Recalibrated linear predictor (B) = -0.5206 + (0.6963 *original fullPIERS model predictor)

6.2.2.1 Performance for recalibrated models

6.2.2.1.1 Discrimination

Neither recalibrated models A or B demonstrated significant changes in the discriminatory performance of the fullPIERS model. Recalibrated model A had an AUROC of 0.82 (95% CI: 0.77-0.87) while recalibrated model B had an AUROC of 0.81 (95% CI: 0.76-0.87), compared to AUROC of 0.81 (95% CI: 0.76-0.87), using the original fullPIERS model.

6.2.2.1.2 Calibration

For both recalibrated models, the intercepts were fixed to an ideal 0 vs 0.3 using the original model (**Figure 6-1**). For the Recalibrated model A, the slope remained at 0.7 while for the Recalibrated model B, the calibration graph became ideal with a slope of 1 versus 0.70.

6.2.2.1.3 Stratification and classification

There was no improvement in the stratification of women into low or high risk categories for both recalibrated models (**Table 6-1**)**.** For recalibrated model A, there minimal movement in the

number of women in each predicted risk groups, except a slight decrease in the lowest risk group (predicted probabilities <1%) and increase in the highest risk group (predicted probabilities ≥ 30%). These movements did not result in any significant change in the classification accuracy although a few more women with adverse outcomes (4 women) would have been correctly identified in the highest risk category.

For recalibrated model B, the stratification performance of the model decreased as there were movements from the lowest and highest risk groups (the extreme risk groups) into the middle risk groups. Classification accuracy at the highest risk group did not significantly change.

6.2.3 Discussion

6.2.3.1 Main findings

Model recalibration is generally recommended to fix poor calibration performance.[21] The recalibration of the fullPIERS model improved the calibration performance of the model but had no effect on the discrimination. The recalibrated model may have corrected for the slight increase in the rates of adverse outcomes between the primary external validation cohort and the fullPIERS development cohort, although this was similar (5% vs 6.6%).

However, stratification did not improve and may have decreased in the recalibration involving both the model intercept and slope. The new generated slope (0.6963) suggests the need for shrinkage of the model predictor coefficients towards zero (0).[21,109] However, this decrease in stratification may suggest that not all the predictor coefficients need to be shrunk or perhaps, some differences in predictor effects in the new population, which might require selective re-estimation of the predictors. Alternatively, the results could also suggest that an important predictor is missing.

6.2.3.2 Comparisons with other findings

Improvements in calibration performance upon recalibration of the model intercept (and slope) have also been reported in several studies.[21,112-116] These studies also reported no change in the discriminatory performance of their models. However, stratification performance of the recalibrated models was not assessed in majority of these studies except for the study by Schuetz *et al*.[114] This study which examined the recalibration of three different models for the prediction of mortality in patients with pneumonia, reported a decrease in classification accuracy in two of the three models. This study also reported that the misclassified patients were more likely to be younger; therefore, it is possible that age is an important predictor in the study by Schuetz *et al*.[114]

6.2.4 Conclusions

Simpler updating methods (recalibration) are recommended[21] and based on the results in this chapter, perhaps only the intercept update should be recommended for recalibration. Ultimately, the recalibration method of choice would depend on what is most useful in making clinical decisions. Since discrimination was not altered, one could decide to choose a logistic recalibration to have a perfect calibration at the expense of risk stratification. Calibration would enable communication of risks to patients although stratification might be more useful for making clinical decisions, especially where outcomes are severe and there are fewer resources. Future studies should try to identify possible reasons for misclassification using the fullPIERS model as this may give insights to what predictors might need re-estimation and any missing predictors that might be useful in the new population.

6.3 Addition of PlGF

6.3.1 Methods

6.3.1.1 Datasets

The PELICAN cohort and the Alere-funded PETRA study, which had available maternal PlGF concentrations measurements, were used. A limited number of cases from the initial development cohort (from BCW Hospital, Canada) also had available PlGF measurements and were included in the cohort. The PELICAN data consisted of a prospective cohort of 625 women admitted with pre-eclampsia in a consultant-led maternity unit in London, United Kingdom from January 2011 to February 2012. The Alere-PETRA data consisted of a prospective cohort of 644 women admitted with pre-eclampsia at twenty-four maternity units in the United States of America (USA) and Canada from May 2010 and November 2011.We restricted our analyses to women with PlGF concentrations measured during hospital admission for pre-eclampsia (or at most 14 days before admission).

6.3.1.2 PlGF measurement

PlGF measurements have been described explicitly elsewhere.[76,83,86] In brief, maternal plasma samples were collected from eligible women during the PELICAN and Alere-PETRA recruitment. The samples were tested for PlGF using the Triage PlGF Test (Alere, San Diego, CA) according to the manufacturer's instructions. Low PlGF was defined as a value <100 pg/mL), while normal PlGF was defined as a value ≥100 pg/mL, pre-classified based on previous studies.[76,83]

6.3.1.3 Definition of pre-eclampsia and adverse outcomes

The same definitions for pre-eclampsia and adverse maternal outcomes as in the model development were used.[24]

An adverse fetal outcome was a composite outcome of one or more of the following: stillbirth, admission into neonatal intensive care unit (NICU) for at least 48 hours, or neonatal death.

6.3.1.4 Statistical analyses

The distribution of patient characteristics of the "Extension" datasets (PELICAN and Alere PETRA) were compared with the development data. Univariable analyses were also carried out comparing characteristics of patients in the extension data with low vs normal PlGF levels.

6.3.1.5 Prediction of adverse maternal outcomes and model extension with PlGF

PlGF measurements were converted to percentiles for normal reference range by gestational age interval.[117] The univariable discriminatory performance of PlGF for predicting adverse maternal outcomes within 48 h of admission was assessed. The ability of PlGF to predict adverse fetal outcomes occurring within 7 days, 14 days and at any time from admission with pre-eclampsia was also assessed.

Before addition of PlGF to the fullPIERS model, the original model performance in this cohort was assessed. For extension of the model, a linear predictor variable was calculated for all the women in the extension dataset using the worst measured model predictor variables within 48 hours of admission as described in section **5.2.1**.[21] A logistic regression model was then fitted with two variables: (i) the linear predictor and (ii) PlGF percentiles. Hence, a new intercept and slope were estimated for the fullPIERS model as well as a regression coefficient for PlGF. The extended model was then used to calculate the predicted probabilities of experiencing an adverse outcome for each woman and its performance was evaluated.

6.3.1.6 Performance measures

The univariate ability of PlGF to predict adverse maternal and fetal outcomes and the performance of the original fullPIERS model were evaluated based on discrimination capacity.

In addition to discrimination, the extended model performance was evaluated based on the Net Reclassification Index (NRI) and integrated discrimination improvement (IDI) (described in **Chapter 3**) to assess the incremental value of the PlGF compared to the performance of the original model in the extension dataset.[21,97,101]

6.3.1.7 Sensitivity analyses

We conducted a sensitivity analyses for the univariate analyses for the ability of PlGF for predicting maternal and fetal outcomes within 48 hours and 7 days of admission respectively, for women admitted with pre-eclampsia before 35 weeks of gestation.

6.3.2 Results

6.3.2.1 Demographics

In total, 600 women (N=541 from the Alere-PETRA cohort, N=50 from PELICAN cohort and N= 9 from the development cohort) who met our inclusion criteria were included in the data. **Table 6-2** presents the characteristics of the women in the Extension cohort. The women in the extension cohort appeared to be younger, more likely to be multiparous and to have an earlier onset of pre-eclampsia compared to the original development cohort. They were also more likely to be smokers, have a higher rate of $MgSO_4$ administration, a shorter admission to delivery interval for women with early onset of pre-eclampsia ($GA<34^{+0}$ weeks), and babies with lower birth weights compared with the development cohort. There were no meaningful differences in the proportions of multiple pregnancies and treatment with antihypertensive medication.

The rates of adverse maternal outcomes were 7.8%, 9.3% and 11% occurring within 48 hours, seven days and at any time respectively. There were 12 cases of stillbirth, 14 cases of neonatal death and 23 cases of admission in neonatal intensive care unit (NICU) for over 48

hours. In total, the rate of adverse fetal outcomes was 5.7%, 6.7% and 7.5% within 7 days, 14 days and at any time from admission.

Patient characteristics according to their PlGF group are shown in Table 6-3**.** PlGF was low in 89.7% (N=537) of the women in the cohort. Women who had an earlier onset of pre-eclampsia, higher blood pressures and babies with lower birth weights were more likely to have low PlGF. Of the women with composite maternal and fetal outcomes at any time, 93.9% (62/66) and 98% (44/45), respectively, had low PlGF.

6.3.2.2 Univariable performance

In the univariable analyses, PlGF showed poor discriminatory performance for the prediction of adverse maternal outcomes within 48 hours of admission (AUROC of 0.60 (95% CI: 0.52-0.670) (**Figure 6-2**) although it performed well in predicting adverse fetal outcomes within 7 days, 14 days and at any time, with AUROCs of 0.76 (95% CI: 0.67-0.85), 0.75 (95% CI; 0.67-0.83), and 0.72 (95% CI: 0.64-0.80), respectively **(Figure 6-3)**.

6.3.2.3 Addition of PlGF on model performance for maternal outcomes

The original fullPIERS model performed poorly in this cohort with an AUROC of 0.67 (95% CI: 0.58-0.76). Extension of the model with PlGF resulted in a beta coefficient of -0.01434 for PlGF percentiles; the equation from the extension of the fullPIERS model with PlGF is shown in **Box 5-2**.

Box 5-2:
Linear predictor for extended model with PlGF = -1.33524 + (0.24854 * original fullPIERS model predictor) + (-0.01434 * PlGF)

Addition of PlGF to the model did not improve the discriminatory performance of the model, AUROC of 0.67 (95% CI; 0.59-0.75) compared with the original fullPIERS model **(Figure 6-2)**.

Using a threshold of ≥10% for the calculated predicted probabilities, there were 8 upward movements of women with adverse maternal outcomes into the high risk category and 2 less cases without adverse maternal outcomes in the highest risk category compared to the classification using the original model **(Table 6-4)**. The overall improvement in sensitivity was 12.8% with a decrease of 10.7% in specificity using the extended model. Therefore, the NRI was 2.1%.

The discrimination slope for the extended model was 0.03 versus 0.09 for the original model (**Figure 6-4**). The IDI was calculated to be -0.06, indicating no improvement in discrimination with the extended model.

6.3.2.4 Prediction performance in GA less than 35 weeks

PlGF also performed well in the prediction of adverse fetal outcomes within 7 days of admission for women with onset of pre-eclampsia less than 35^{+0} weeks of gestation (AUROC of 0.79 (95% CI: 0.72-0.87)) but poorly for the prediction of adverse maternal outcomes within 48 hours of admission (AUROC of 0.52 (95% CI: 0.43-0.61)) **(Figure 6-5)**.

6.3.3 Discussion

6.3.3.1 Main findings

The findings showed a poor performance of the fullPIERS model in this cohort. We hypothesize that the reduced performance of the model in this cohort was due to the case-mix differences between the extension cohort and the development cohort which may have resulted from differences in disease onset, demographics and management of pre-eclampsia between the cohorts.[103,104,111] The extension data was mostly made up of the Alere-PETRA cohort which consist of women admitted in the USA. As previously described in **Chapter 4**, the women in this cohort were less likely to be managed expectantly compared with the women included in the

development cohort (development centres were chosen because of a policy of expectant management remote from term), as evidenced in the shorter admission-to-delivery interval in the extension cohort. Such interventionist management does not allow the natural trajectory of the disease since women are more likely to be delivered earlier, thus prediction of adverse outcomes is less accurate. The addition of PlGF to the model did not improve its prognostic ability for adverse maternal outcomes. In general, PlGF had a poor prognostic ability for the prediction of adverse maternal outcomes for women admitted with pre-eclampsia, whether as an independent marker or upon addition to the fullPIERS model. However, PlGF appeared to discriminate well for women with composite fetal outcomes, especially for women admitted before 35 weeks of gestation.

6.3.3.2 Strengths and limitations

This is the first study of which we are aware to assess the addition of PlGF to the fullPIERS model for the prediction of adverse maternal outcomes. The findings from this study are important to direct the use of the model in addition with PlGF for HDP management. Our study also investigated the prognostic value of PlGF alone for severe maternal outcomes and other severe fetal outcomes including stillbirths and neonatal deaths for women admitted with pre-eclampsia which are common.

There were several limitations in this study. We had limited power due to the small sample size and relatively small sample size of adverse outcomes. As a result, we were unable to re-estimating the fullPIERS model coefficients when extending the model with PlGF to prevent overfitting of the model in the data set.

Another limitation was the timing of PlGF measurement. Ideally for the fullPIERS model, PlGF measurement should be carried out within 48h of admission but due to much earlier assessment,

PlGF measures had to be carried forward in this study. Although the test measurement was restricted to 14 days, pre-eclampsia is a progressive disease and the levels of PlGF can change quickly within a short period.[13,14] Thus, this could have affected the prognostic performance of PlGF.

6.3.3.3 Comparisons with other findings

As shown in the review for PlGF **in Chapter 2**, a few studies reported strong prognosis for the prediction of combined adverse and maternal outcomes form HDPs. The studies by Ghosh *et al*[57] and Leaños-Miranda *et al*[58] also reported increased risk of adverse maternal outcomes for women with pre-eclampsia with low PlGF, although the data from their studies did not show any strong prognostic value of PlGF to either rule in or rule out adverse maternal outcomes. Therefore, it is possible that the positive results reported by studies for the prediction of combined maternal and fetal outcomes are being driven mostly by the predictive performance of PlGF for the fetal outcome components. Of note, majority of these studies combined PlGF with sFlt-1 expressed as a ratio.[61;80;84-85]

The findings from this assessment support results from other studies whereby PlGF has been reported to be a good predictor of composite perinatal outcomes including stillbirths, neonatal deaths and significant neonatal morbidities of babies born to women with suspected or confirmed pre-eclampsia.[80-85] In this study, almost 90% of the women had low PlGF measures. The high rate of low PlGF concentrations in our cohort is consistent with the growing evidence that reduction in the angiogenic marker is closely related to the development of pre-eclampsia. In addition, similar to a previous study by our team, women with low PlGF (<100 pg/mL) were more likely to have higher BP and experience worse neonatal outcomes. Although lower PlGF might be predictive of pre-eclampsia and associated with worse clinical measures, it was not

predictive of adverse maternal outcomes in this study. The mechanism for its involvement in pre-eclampsia remains unclear; however, it was predictive of fetal outcomes in our study. Studies have reported that PlGF is most useful for prediction before 35 weeks of gestation as the marker begins to fall after then, even in normal pregnancies. We also observed this trend of improvement in the discriminatory performance for the prediction of neonatal outcomes for women with gestational age of less than 35 weeks.

6.3.4 Implications

The addition of PlGF did not improve the performance of the fullPIERS model in predicting adverse maternal outcomes. While the usefulness of PlGF for predicting adverse maternal outcomes is uncertain, PlGF could be considered as a potential predictor of adverse fetal outcomes for women with pre-eclampsia. PlGF could aid in early detection of pregnancies complicated by pre-eclampsia that will experience severe fetal outcomes which might be prevented by timely and more appropriate interventions such as closer monitoring, administration of corticosteroids, earlier delivery, and transfer to higher care units. Larger, better-designed cohort studies on women admitted with pre-eclampsia and managed expectantly might be beneficial to determine if the differences in timing and threshold of PlGF measurements can improve its prognostic ability.

6.4 Summary

Recalibration methods corrected for miscalibration of the fullPIERS model in the datasets although stratification performance. PlGF did not add any incremental value to the fullPIERS model but appeared useful for the prediction of adverse fetal outcomes. When applying the model to another setting, intercept updating may be considered.

Table 6-1 Risk stratification table of the original and recalibrated model in the Primary External Validation dataset

	Original fullPIERS model		Updated model intercept only		Updated model intercept and slope	
Prediction score range	**Total N women in range (%) (N=1504)**	**N women with outcome (%) (N=99)**	**Total N women in range (%) (N=1504)**	**N women with outcome (%) (N=99)**	**Total N women in range (%) (N=1504)**	**N women with outcome (%) (N=99)**
<1.0%	459 (30.5%)	8 (1.7%)	364 (24.2%)	5 (1.4%)	76 (5.1%)	2 (2.6%)
1·0-2·4%	498 (33.1%)	14 (2.8%)	459 (30.5%)	12 (2.6%)	407 (27.1%)	7 (1.7%)
2·5-4·9%	287 (19.1%)	13 (4.5%)	316 (21.0%)	10 (3.2%)	518 (34.4%)	14 (2.7%)
5·0-9·9%	117 (7.8%)	16 (13.7%)	183 (12.2%)	16 (8.7%)	317 (21.1%)	20 (6.3%)
10·0-29·9%	77 (5.1%)	12 (15.6%)	103 (6.9%)	16 (15.5%)	128 (8.5%)	24 (18.8%)
≥30·0%	66 (4.4%)	36 (54.5%)	79 (5.3%)	40 (50.6%)	58 (3.9%)	32 (55.2%)

Table 6-2 Maternal characteristics for the Extension data (Alere PETRA, PELICAN and BCW)

Characteristics	fullPIERS cohort (development) (2,023 women)	PlGF cohort (Extension) (600 women)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS		
Maternal age at EDD (yr)	31 [27, 36]	30 [24, 34]
Parity ≥1	581 (28·7%)	242 (40.3%)
Gestational age at eligibility (wk)**	36 [33, 38·3]	33.4 [29.9, 36]
Gestational age at eligibility <35 weeks, N	790 (39.1%)	380 (63.3%)
Multiple pregnancy	192 (9·5%)	47 (7.8%)
Smoking in this pregnancy	249 (12·3%)	120 (20.0%)
CLINICAL MEASURES		
Systolic BP (mm Hg)	160 [150, 176]	144 [135, 157]
Diastolic BP (mm Hg)	102 [98, 110]	87 [78, 95]
Chest pain/dyspnoea**	90 (4.4%)	12 (2.0%)
Uric acid	376 [320, 427]	369 [307, 434]
Lowest platelet count ($\times 10^9$ per L)**	192 [150, 242]	197 [152, 243]
Highest AST/ALT (U/L)**	28 [21, 41]	24 [18, 35]
Creatinine	67 [58, 77]	62 [53, 71]
INTERVENTIONS DURING ADMISSION		
Corticosteroids	550 (27·2%)	163 (27.2%)
Corticosteroids, GA onset <35	473/790 (59.9%)	151/380 (39.7%)
Antihypertensive therapy	1381 (68·3%)	435 (72.5%)
$MgSO_4$	690 (34·1%)	425 (70.8%)
PREGNANCY OUTCOMES		
Admission-To-Delivery Interval (Days)	2 [1, 5]	2 [1, 5]
Admission-To-Delivery Interval, <34^{+0} Weeks (Days)	4 [2, 14]	3 [1, 6]
Gestational age at delivery (wk)	36.9 [34·1, 38·6]	34.1 [30.6, 36.7]
Birth weight	2141 [1441, 2807]	1991 [1204, 2730]
Stillbirth	20 (1.0%)	10 (1.7%)
Neonatal death	26 (1·3%)	14 (2.3%)
NICU admission ≥48hrs	-	23 (3.8%)
MATERNAL OUTCOME (N women)		
Within 48h	106 (5.2%)	47 (7.8%)
Within 7 days	203 (10.0%)	56 (9.3%)
At anytime	261 (12.9%)	66 (11.0%)

*** Variables included in the model*

Table 6-3 PlGF levels in the Extension data

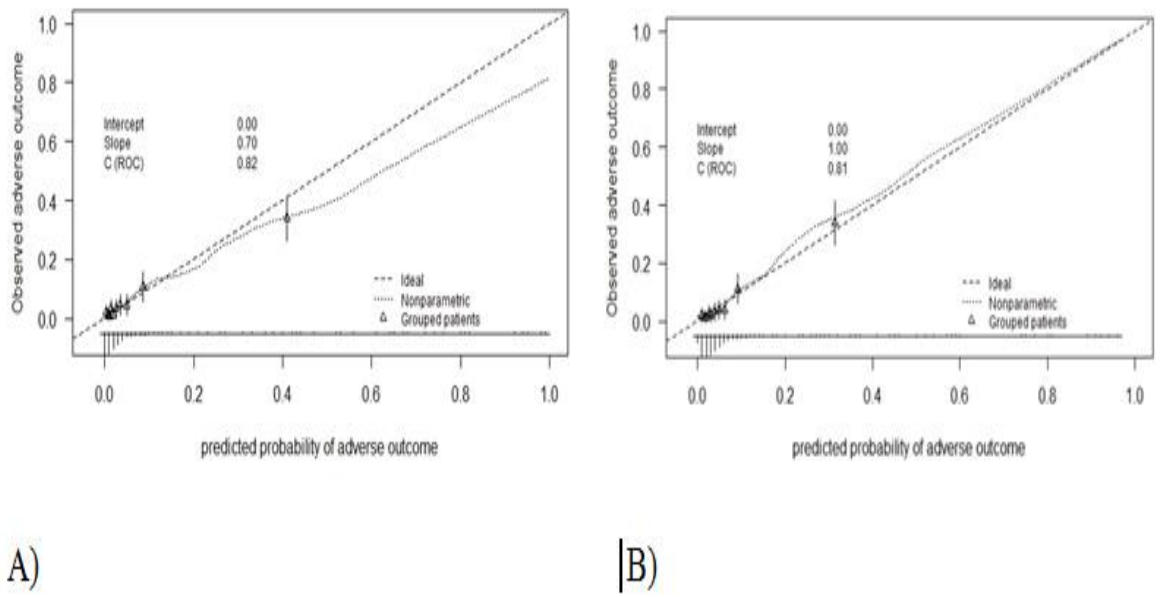
Characteristics	Low PlGF (<100 pg/ml), n (%) or median (IQR), (N=537)	Low PlGF (≥100 pg/ml), n (%) or median (IQR), (N=63)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS		
Maternal age at EDD (yr)	30 [24, 34]	36 [26, 35]
Parity ≥1	212 (39.5%)	30 (47.6%)
Gestational age at eligibility (wk)**	32.9 [29.6, 35.9]	35.4 [33.4, 36.9]
Gestational age at eligibility <34 weeks, N	304 (56.6%)	22 (34.9%)
Gestational age at eligibility <35 weeks, N	352 (65.6%)	28 (44.4%)
Multiple pregnancy	43 (8.0%)	4 (6.4%)
Smoking in this pregnancy	103 (19.2%)	17 (27.0%)
CLINICAL MEASURES		
Systolic BP (mm Hg)	145 [136, 158]	138 [127, 150]
Diastolic BP (mm Hg)	87 [78, 95]	80 [75, 92]
Chest pain/dyspnoea**	11 (4.4%)	1 (1.6%)
Uric acid	375 [315, 435]	317 [265, 365]
Lowest platelet count ($\times 10^9$ per L)**	196 [152, 243]	225 [159, 249]
Highest AST/ALT (U/L)**	24 [19, 37]	18 [16, 24]
Creatinine	62 [53, 71]	53 [44, 63]
INTERVENTIONS DURING ADMISSION		
Corticosteroids	155 (28.9%)	8 (12.7%)
Corticosteroids, GA onset <35	143/352 (59.9%)	8/28 (39.7%)
Antihypertensive therapy	394 (68·3%)	41 (72.5%)
$MgSO_4$	391 (72.8%)	34 (54.0%)
PREGNANCY OUTCOMES		
Admission-To-Delivery Interval (Days)	2 [1, 4]	2 [1, 11]
Gestational age at delivery (wk)	33.7 [30·1, 36.7]	36.1 [34.7, 37.3]
Cesarean delivery (Y)	362 (67.4%)	29 (46.0%)
Birth weight (g)	1820 [1179, 2660]	2750 [2335, 3135]
Stillbirth	10 (1.9%)	0
Neonatal death	14 (2.6%)	0
NICU admission ≥48hrs	22 (4.1%)	1 (1.6%)
Composite fetal (7 days)	30 (5.6%)	0
Composite fetal (14 days)	37 (6.9%)	0
Composite fetal (anytime)	44 (8.2%)	1 (1.6%)
MATERNAL OUTCOME (N women)		
Within 48h	44 (8.2%)	3 (4.8%)
Within 7 days	53 (9.9%)	3 (4.8%)

Characteristics	Low PlGF (<100 pg/ml), n (%) or median (IQR), (N=537)	Low PlGF (≥100 pg/ml), n (%) or median (IQR), (N=63)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS		
At anytime	62 (11.6%)	4 (6.4%)

** *Variables included in the model*

Table 6-4 Reclassification table

		Model with PlGF		
	Women with events, %	**0 to 9**	**≥ 10**	Total
Model without PlGF	0 to 9	28	8	36
	≥ 10	2	9	11
	Total	30	17	47
	Women without events, %			
	0 to 9	464	61	525
	≥ 10	2	26	28
	Total	466	87	553



A) B)

Figure 6-1 Validation graphs for A) Recalibrated model intercept, and B) Recalibrated model intercept and Slope

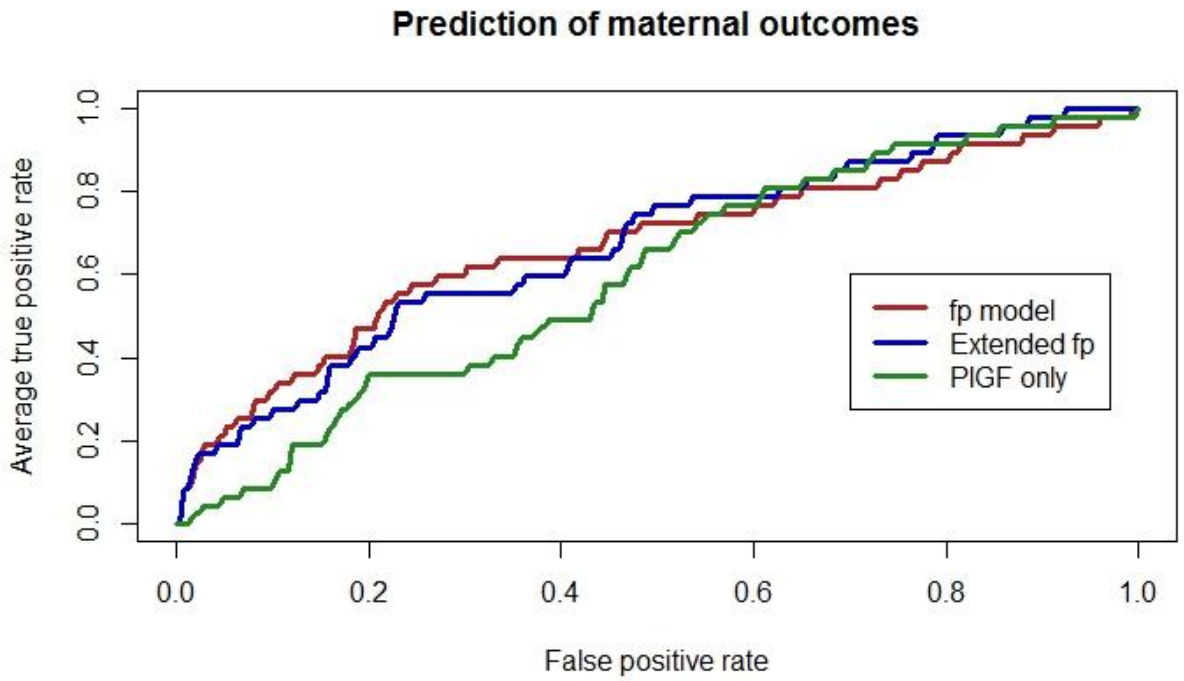


Figure 6-2 Prediction of adverse maternal outcomes using (i) only PlGF (ii) original fullPIERS (fp) model and (iii) the extended fullPIERS model (AUROC of 0.60, 0.67 and 0.67 respectively)

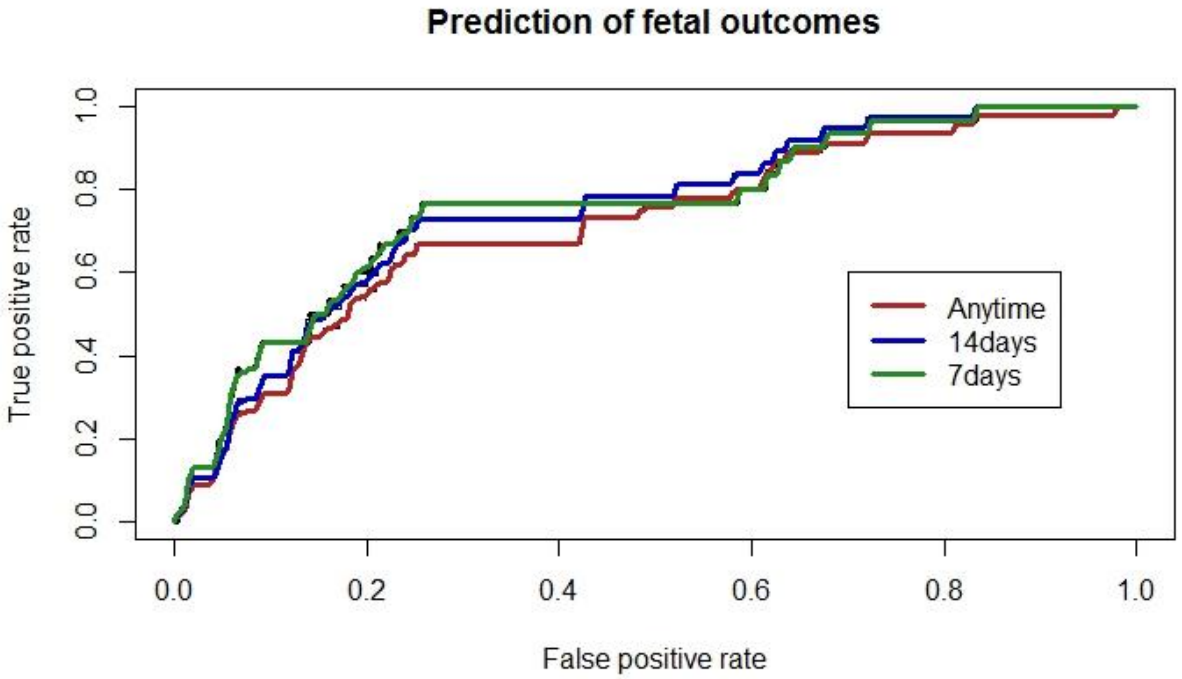
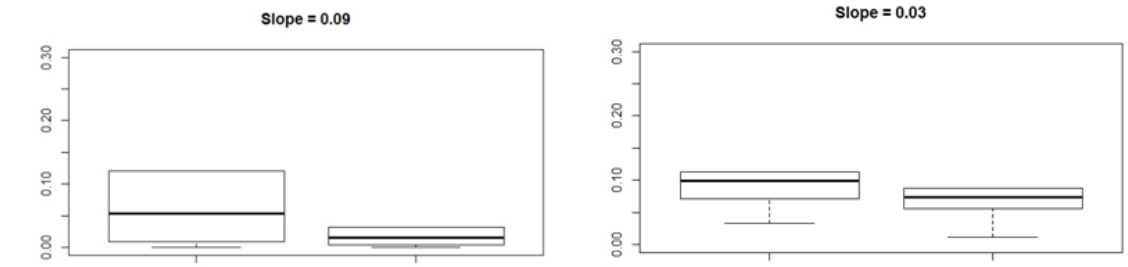


Figure 6-3 Prediction of fetal outcomes occurring within 7 days, 14 days and any time from admission with pre-eclampsia with PlGF (AUROC of 0.76, 0.75 and 0.72 respectively)



(A) B)

Figure 6-4 Discrimination slopes for (A) Original model (without PlGF) and (B) Extended model with PlGF (IDI = -0.06)

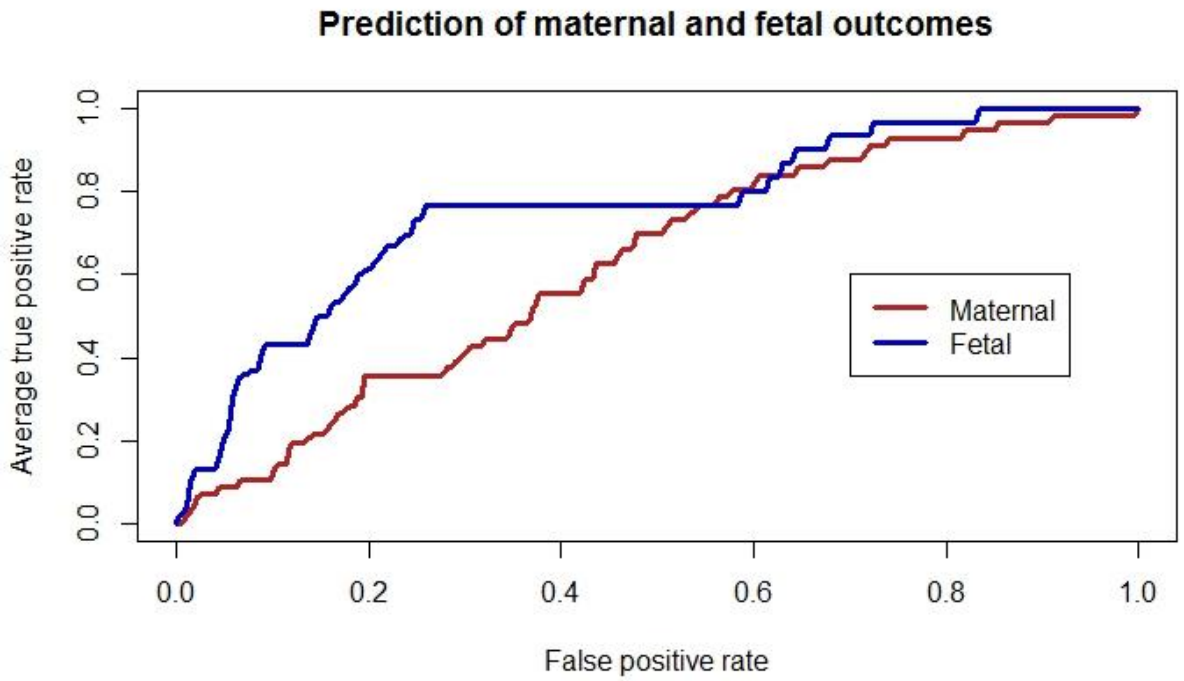


Figure 6-5 Prediction of adverse maternal and fetal outcomes for women admitted before 35 weeks of gestation (AUROC of 0.52 and 0.79 respectively)

Chapter 7: Factors Influencing Model Performance

7.1 Background

The performance of a risk prediction model is generally lower when assessed in a population other than the development cohort.[21,23,25] One of the reasons for this is because of overfitting of the model in the development cohort.[21,27] The reasons for overfitting include small effective sample size, too many predictors included in the model and poor selection methods of model variables such as backward selection based on *p*-values.[21, 110] Generally, assessing for overfitting using methods such as bootstrapping gives an idea of how well the model will perform in the future. [21,110] Over-optimism of the model performance due to overfitting is quantified and used to calculate the corrected internally validated performance of the model. Shrinking of the model regression coefficients is also recommended to improve the model performance in future subjects.[21]

Regardless of how well the statistical methods used to develop a model are, the model might not perform as well as when tested in external populations because of differences between the development cohort and future/external cohorts (or validation cohort).[21,107,108] These differences could be due to different distribution of outcomes or predictor variables between the development and validation populations related to the outcome (case-mix differences), or might also include missed or unmeasured predictors.[103,104] Therefore, case-mix differences include using a validation cohort with more or less severe group of patients, or more or less heterogeneous population, or higher or lower rates of outcomes, than used in the development. Another reason for a change in model performance is that the predictor coefficients (or effects) are different in the new population.[21,103] This often occurs as a result of overfitting, however, there are instances where the validation populations might be truly different from the

development population. For example, a model development study found different predictor effects of the same variables for predicting heart disease in different ethnic groups and races. In such cases, adding ethnicity as a variable or intervention might correct for the effect. Differences in predictor effects can also occur if the new population uses or includes new definitions for the outcomes or predictors.

In the previous **Chapters (4-6)**, these reasons were proposed to be the cause of the observed model performances in the various validation cohorts. However, it will be useful to explore these reasons for performance change in more detail to give more conclusive information on the model performance in the cohorts. Comparisons of the distributions of patient characteristics between the development and validation datasets have already been presented to highlight possible sources of case-mix differences. In this chapter, the extent to which case-mix differences or/and differences in predictor effects in some of the cohorts affected the fullPIERS model performance are further examined.[21,103,104]

In addition, possible factors contributing to misclassification of women into relevant risk groups, particularly false negative, are examined.

7.2 Methods

7.2.1 Linear predictor

To examine the case-mix differences regarding the severity of pre-eclampsia and heterogeneity of the population, the linear predictor was calculated for each woman, as described in **Chapter 5**, in some of the datasets: temporal, primary external, Dutch PETRA, PREP and miniPIERS, and for the fullPIERS cohort. These datasets were used to demonstrate the model effects for the different spectrum of patients in the validation studies. The linear predictor (rather than the calculated predicted probabilities) is preferred because it allows for easy interpretation since it is

on the logit transformation scale.[103] An increase in the mean of the linear predictor (compared to the original model) signifies more severity, while an increase in the standard deviation of the linear predictor signifies more case-mix heterogeneity.[21,103]

7.2.2 Reference values

Reference values were also calculated in the same datasets: temporal, primary external, Dutch PETRA, PREP and miniPIERS. The use of 'reference values' quantifies what the model predictions should be, if the original regression coefficients are valid for the new (validation) dataset; this takes into account the distribution of the predictors by simulating random outcomes for each patient while assuming that the prediction model is correct.[21] Thus, the reference values help to answer whether the differences in the model performance are due to differences in case-mix, regression coefficients, or both. If the reference value is similar to the model performance in the development data, then difference in model external validation performance is more likely to be explained by differences in regression coefficients; if the reference value is different, then the difference in model performance can be explained by case-mix differences.[21] The external model performance and the reference values are also compared to interpret possible predictor effects. Using this approach, 1000 repetitions were simulated per patient to obtain a stable estimate of the reference values.

7.2.3 Predictor coefficients

The fullPIERS model variables were refit in the datasets to generate new regression coefficients which were compared to the original model coefficients. This was not used to generate a new model performance but to provide insight as to how the predictor effect may have differed in the datasets.

7.2.4 Classification accuracy

Using the calculated probability of ≥ 30% as a threshold for 'high risk', the characteristics of the women who were correctly identified (true-positives) versus those were not (false negatives) were compared to help in identify possible additional factors associated with adverse maternal outcomes.

7.3 Results

There was no significant difference in the mean and the standard deviation of the linear predictor for the temporal and primary external cohorts compared with the fullPIERS cohort (**Table 7-1**); however, the mean was higher for the Dutch PETRA cohort (-2.921 in Dutch PETRA vs -3.956 in fullPIERS), indicating a more severe case-mix of patients (for observed predictors) but lower in the PREP and miniPIERS cohort indicating a less severe case-mix of patients. In addition, both the Dutch PETRA and the miniPIERS cohort had higher case-mix heterogeneity (i.e., higher standard deviations), with the Dutch PETRA having the highest standard deviation.
The reference values for the temporal and primary external cohorts were similar to the model development value also suggesting that the slight decrease on external assessment (reduced performance in external i.e. AUROC of 0.81 vs 0.88) is not explained by case-mix differences. Rather, the non-significantly poorer external validity is likely attributable to differences in regression coefficients between the settings. In the case of the Dutch PETRA cohort, the reference value was different from the model development value indicating case-mix differences. The better performance of the model in this dataset compared with the reference value (0.97 vs 0.92) also suggests stronger predictive effects in the validation sample. For the PREP and miniPIERS cohort, the external performance can also be partly attributed to case-mix differences

(reference value is different from model development for the miniPIERS) and partly to different predictor effects (different reference values compared to external performance).

In the refit model (**Table 7-2**), all the regression coefficients generally appear bigger for the Dutch PETRA cohort. In some of the cohorts, there was also a reversal in the direction of effects for some variables, e.g. gestational age, which might suggest the need for re-estimation. In comparing the true positive cases with the false negative cases, the only main differences found were that women who were correctly identified were more likely to have right upper quadrant pain, headache and more likely to smoke **(Table 7-3)**.

7.4 Discussion

7.4.1 Main findings

This chapter examined possible reason for the changes in model performance found in the different validation cohorts included in this thesis. As hypothesized in the individual chapters, the temporal and primary external cohorts did not have any meaningful case-mix differences in the populations compared with the development cohort. In contrast, the cohorts used for the domain validation had substantial case-mix differences. The Dutch PETRA cohort showed both differences in the severity and heterogeneity case-mix as well as in predictor coefficients, which appeared stronger. This supports the hypothesis that the inclusion of more women with HELLP syndrome and other severe cases may have resulted in the improved AUROC performance. Comparison of the true positive and false negative cases using the temporal data indicated that women with right upper quadrant pain, headache and who were more likely to smoke, were more likely to be correctly classified as high risk patients. These variables could be examined as potential predictors in the future.

The PREP and miniPIERS cohorts showed a case-mix of less severity based on their linear predictors. However, this is not to say that the women in these cohorts were less sick compared with the Dutch PETRA or the development cohorts since the linear predictors are only calculated based on the predictors in the model. As shown in the results in **Chapter 6**, the rate of adverse outcomes was higher for the miniPIERS cohort. Women in LMICs are more likely to have other comorbidities which may worsen their rates of adverse outcomes, thus it is likely that the effect of the fullPIERS predictors are less in the population and there might be a need for other variables to be included in the model as either predictors or as interaction terms to identify high risk women more accurately.

7.5 Conclusion

The reduced fullPIERS model performance in the domain validation datasets can be attributed to both case-mix differences and differences in predictor effects. However, there may have been some "missed" predictors which need to be included, especially in LMICs. Our results also suggest that women with HELLP syndrome who will go on to experience an adverse outcome are more likely to be correctly identified by the model.

Table 7-1 Evaluating reasons for model performance

Performance measures	fullPIERS	BCW	Primary External	Dutch PETRA	PREP	miniPIERS
Mean lp (SD)	-3.956 (1.52)	-3.867 (1.60)	-3.903 (1.55)	-2.921 (2.03)	-4.640 (1.45)	-4.128 (2.48)
t-test for lp		0.0937	0.3130	<0.0001*	<0.0001*	<0.0001*
AUROC	0.88	0.82	0.81	0.97	0.73	0.77
Reference value	-	0.88	0.88	0.92	0.87	0.90

SD – Standard deviation, lp – linear predictor, *significant *p*-values

Table 7-2 Refit model coefficients

Variables	fullPIERS	BCW	Primary External	Dutch PETRA	PREP	miniPIERS
Gestational age at Eligibility	–5·41×10–²	-4.700×10^{-3}	4.493×10^{-3}	1.81 ×10^{1}	-1.393×10^{-2}	1.485×10^{-2}
Chest pain or dyspnoea	1·23	9.345×10^{-1}	1.000*	1.985×10^{2}	1.958*	1.275*
Creatinine	–2·71×10–²	3.606×10^{-3}	-5.413×10^{-3}	-4.954	-7.281×10^{-2}	-3.359×10^{-4}
Platelets	2·07×10–¹	1.215×10^{-2}	3.492×10^{-2}	5.968×10^{1}	1.773×10^{-1}	3.065×10^{-1}*
Platelets²	4·00×10–⁵	6.337×10^{-5}	5.650×10^{-5}*	1.258×10^{-2}	5.022×10^{-5}	2.687×10^{-5}
AST	1·01×10–²	1.497×10^{-3}	1.312×10^{-3}	5.825×10^{-1}	1.658×10^{-2}	7.580×10^{-3}*
AST²	–3·05×10–⁶	1.305×10^{-6}	9.569×10^{-7}	8.703×10^{-4}	-2.396×10^{-5}	-3.380×10^{-6}*
Creatinine×platelet	2·50×10–⁴	1.449×10^{-4}	1.583×10^{-4}*	-5.326×10^{-3}	4.757×10^{-4}	6.044×10^{-5}
Platelet ×AST	–6·99×10–⁵	-3.785×10^{-6}	-4.324×10^{-4}	-2.683×10^{-2}	-8.469×10^{-6}	-2.284×10^{-5}

Table 7-3 True Positive vs False Negative for Primary external datasets at probability ≥ 30%

Characteristics	True Positive (36 women)	False Negative (51 women)
DEMOGRAPHICS & PREGNANCY CHARATERISTICS		
Maternal age at EDD (yr)	35 [31, 38]	35 [31, 37]
Parity ≥1	7 (19.4%)	8 (15.7%)
Multiple pregnancy	6 (16.7%)	8 (15.7%)
Smoking in this pregnancy	5 (13.9%)	1 (2.0%)
CLINICAL MEASURES		
Systolic BP (mm Hg)	170 [154, 184]	166 [154, 177]
Diastolic BP (mm Hg)	100 [93, 109]	100 [93, 108]
Uric acid	441 [365, 500]	420 [372, 485]
Nausea	11 (30.6%)	11 (21.6%)
Headache	18 (50.0%)	19 (37.3%)
Visual	4 (11.1%)	7 (13.7%)
Right Upper Quadrant pain	15 (41.7%)	4 (7.8%)
Abdominal pain and bleeding	12 (33.3%)	23 (45.1%)
INTERVENTIONS DURING ADMISSION		
Corticosteroids	11 (30.6%)	16 (31.4%)
Antihypertensive therapy	33 (91.7%)	41 (80.4%)
PREGNANCY OUTCOMES		
Birth weight	2475 [1879, 3142]	2410 [1688, 3050]
Male	13 (36.1%)	28 (54.9%)

Chapter 8: Conclusion

The objective of this research was to assess the external validity of the fullPIERS risk prediction model for the prognosis of women with pre-eclampsia and the benefit of adding PlGF to the model.

In the preceding chapters, the model was applied to various external cohorts and the performance evaluated using both traditional measures of AUROC and calibration as well as novel methods such as NRI and IDI for assessing the incremental value of PlGF. The findings from this thesis suggest that the model is valid for the prediction of adverse maternal outcomes in the temporal sites and external sites in related patients (**Table 8-1**). The model still provided valid but less accurate predictions for early-onset pre-eclampsia, and women admitted with other HDPs in LMICs although recalibration (of the intercept). More extensive updating methods may be required for further improved performance. In settings where there is less expectant management, further research is required before the model can be used to guide the management of women with pre-eclampsia.

8.1 Implications of research findings

The ability to predict which women will suffer an adverse outcome within 48 hours could improve care of the woman by guiding transfer to higher care units, administration of corticosteroids and $MgSO_4$, and planning for delivery.

Overall, the fullPIERS model showed good discriminatory performance (AUROC) and acceptable calibration in all the validation cohorts, except in the Alere-PETRA cohort. We hypothesize that the less common use of expectant management in this setting may explain this latter finding. Hopefully, the results from this study can also direct care in settings like the Alere-PETRA to reduce the rates of induced delivery from HDPs, which might not be necessary. Our

findings suggest that the model can be used in similar settings although model updating should be carried out periodically, to maintain the observed performance and ensure changes in clinical management have not impacted its relevance. The results from this study may also be useful for supporting use of expectant management, thereby reducing the rates unnecessary preterm delivery from HDPs.

The fullPIERS model showed the best statistical validity in the temporal and primary external cohort (i.e. it passed statistical thresholds for the performance). However, sometimes clinical validity is more important (i.e. a model that performs satisfactorily in a new setting). For example, the model may be more beneficial for women in LMICs settings than HICs, since women in such settings are more likely to have worse outcomes but also have lower resources. Therefore, early and accurate identification of a high-risk woman in such places could be both effectively life-saving and cost-saving by aiding in reducing adverse outcomes, using fewer resources.

One way to implement the model into practice is by adaptation into mobile applications (apps); this will be useful given the complexity of the fullPIERS logistic equations. This has already been demonstrated in a trial using the miniPIERS model. As an example of directed use, in a clinical setting, laboratory results for a woman admitted with HDPs at less than 37 weeks can be requested for and the relevant fullPIERS variables can be entered in mobile app to calculate her risk of having an adverse event within 48 hours. Delivery should be considered for a woman having ≥30% risk, as demonstrated in the validation studies. For a woman with predicted probability between 10-29%, closer monitoring should be ensured and planning for delivery and other necessary treatments such as $MgSO_4$ should be made available. Woman with < 5% may be discharged home based on the fullPIERS negative predictive value and asked to

revisit for reassessment within a couple of days. The fullPIERS model may also be useful for guiding care to avoid excessive IV fluid that could result in pulmonary oedema and other respiratory complications. For example, diuresis may be initiated for a woman with a higher predicted probability (~20-29%) to maintain appropriate fluid balance. Certainly, decisions on delivery for higher risk women would also be made in association with the clinician's perception or judgment of the woman's condition but the model would be a useful guiding tool, especially in cases of uncertainty. Of note, risk tolerance would also be dependent on gestational age and availability of resources in the healthcare setting. Therefore, a high predicted probability, e.g. 20%, at earlier gestational ages (≤ 28 weeks) where the benefit of expectant management is higher for the neonate, will be considered differently by the healthcare provider compared with the same probability at 36 weeks of gestation. Thus, the balance of risk, benefit, and the capacity to manage complications, while hard to quantify, is essential and at the heart of clinical decisions, to be guided by the fullPIERS model.

Adjusting the thresholds in the classification table can be used to improve the sensitivity of the model at the expense of specificity or vice versa, depending on what is considered most desirable in the clinical setting.

8.2 Strengths and limitations

In this thesis, the fullPIERS model was assessed in a rigorous manner to provide information on its ability to guide clinical care and potential use in clinical practice. The recommended hierarchy for model validation studies was used; firstly, assessing the model performance in the most related populations (in the primary and temporal validations in **Chapter 4**) and then in broader domains (assessed in early-onset pre-eclampsia, other HDPs and in LMICs). These steps provided insights into how the model performance changes with differences in populations.

The combination of different cohorts increased the robustness and generalizability of the results from this study. The data used for the primary external validation were collected by data collectors trained for the PIERS project; the data were cleaned to identify any data errors and outliers were checked for accuracy. In the BCW cohort, re-abstraction of information from patient records and data re-entry were done in over 5% of the data, to assess the potential for data abstraction errors. For the other cohort sites, data were also collected using good methods related to their independent studies.

An additional strength of this thesis is the exploration of model updating methods to improve the observed performance in validation cohorts. Although, addition of PlGF did not show any benefits, it is important to highlight this to prevent predictions of maternal complications based on the combination unless a better quality study shows improved performance with fullPIERS and PlGF. However, since PlGF demonstrated a potential for predicting preterm delivery, perhaps it can be used as a complimentary aid to guide management for corticosteroid administration and planning for NICU admission.

The main limitation of the thesis was the sizeable amount of missing data in some of the datasets used in this study. Multiple imputation was used to generate plausible values for the missing observations, which is recommended for data missing at random. However, it is possible that some of the data were not missing at random and this may have affected the results and interpretation, especially in the broader cohorts used (from Oxford). To reduce the risk of bias resulting from missing data, the results and conclusions for the primary external validation study were based on the most complete cohorts.

Another limitation of this thesis is that we were unable to correct for the possible differences in predictor effects hypothesized in Chapter 8. ‘Correcting’ the model regression coefficient in each

data could result in even more biased estimates because the true predictor regression coefficients for the population may not be obtainable in the data.[18] Thus, re-estimating model coefficients would require large, randomly sampled dataset that is representative of the true population and events, to prevent overfitting.

8.3 Future research directions

Future research directions could include randomized controlled trials, followed by an implementation or impact study to evaluate the utility of applying the model outside the research setting.21 Prior to implementation, clinicians will need to be reassured that the use of the model will facilitate decision-making by clinicians, women and their families, and, in settings in which adverse outcomes are common, improve clinical outcomes. Such a trial would focus on process of care indicators.

The outcomes that should be evaluated in the impact study should include the utility of the model by care providers, measure how the model affects or changes their practice, and any changes in the rates of adverse outcomes. For such a study, it is important to have an intervention arm where the model is being used and a control arm, to allow for effective comparison of outcomes of interest. The implementation process should provide clear directions on which clinical actions (i.e., delivery, monitoring, expectant management) are recommended for different ranges of predicted probabilities. This would involve knowledge translation and communication with care providers. Discussion with clinicians is very important to facilitate the use of the model and to obtain useful feedback on how the model can be used. An informal discussion with some of the clinicians at BCW Hospital suggested that some clinicians would be willing to use the model if it were easily available and if they understood what the predicted

probabilities imply. An implementation study giving clear directions on how the model can be used (such as the example mentioned earlier) might help.

Another important aspect is that the model to clearly communicate is that the model is to be used only for women who have at least a type of HDP, preferably pre-eclampsia. One of the clinicians had expressed dissatisfaction on the model's inability to identify a fast-brewing patient who had high liver enzymes (not HELLP syndrome). The patient's enzymes (AST and ALT) continued to increase rapidly and so she was delivered based on fear of liver rupture or dysfunction. However, this woman had no hypertension or any form of HDPs; therefore, it is important to emphasize that the model should not be used in non-HDP cases, as this was not the population it was designed for.

A possible question based on the thesis result is "How often should a patient be re-assessed?" As shown in the Chapters 4 and 6, the model works best for predicting adverse maternal outcomes within 48 hours and can also be used for prediction within 7 days, although with less accuracy. Therefore, a woman who has not yet been delivered should be re-assessed at most within 7 days. An important next step to consider is dynamic modelling because it takes into accounts the previous predictor measurements rather than treating them as independent at each repeat assessment. This is essential because the changes in the levels of predictor e.g. AST are likely inter-dependent.

Another question is "how often should the model be updated?" There are no set rules for when a model should be updated. However, it is important to reassess its performances frequently to ensure that the model predictions are still valid, given ongoing evolution of clinical management, and update the model if there are changes observed in its performance. Another trigger for model updating will be the discovery of new biomarkers as stated in Chapter 5. If a

new and cost-effective marker with promising results becomes available, then the incremental value of the marker should be evaluated.

Finally, shrinkage and selective re-estimation of the model parameters using a combined cohort should be considered. The fullPIERS model showed minimal optimism of 0.02, thus its coefficients were not shrunken. Regardless of how low the optimism of a model is, shrinkage has been highly recommended to improve performance on external validation.18

8.4 Intellectual contribution to knowledge and Conclusion

This thesis include many original contributions that I believe would contribute intellectually to knowledge in the area of the development, validation and use of clinical prediction models, and management of women with HDPs. Firstly, I have conducted two thorough systematic reviews presented in **Chapter 2**; one of which focuses on the synthesis of available prediction tests for adverse maternal outcomes resulting from pre-eclampsia and the other on the use of PlGF as a predictor of adverse outcomes from HDPs. The findings from these reviews are important in showcasing potentially useful tests for prognosis in HDPs. I also applied new measures when evaluating the model performance and recalibrating the model, to broaden its interpretation and ways it can be implemented for clinical used. Furthermore, the process of acquisition and sharing of data used in this project from vast groups of research teams globally enabled me to build a strong network of collaborators, who I hope to continue to work with to develop, improve and implement strategies to improve maternal health worldwide.

To conclude, the external validation process is critical for establishing the usefulness of a risk prediction model in the real world setting. Many prediction models are developed that could be of potential use but do not get implemented for use because of lack of external validity. The fullPIERS model has shown external validity for predicting adverse maternal outcomes in

women with pre-eclampsia and other HDPs. It is my hope that the findings from this thesis will contribute to efforts to implement fullPIERS for reducing the global burden of maternal morbidity and mortality through early identification and adequate management of women at the most risk of encountering severe complications from HDPs.

Table 8-1 Summary of model performance

	fullPIERS	Temporal	Primary External	Early onset	miniPIERS	Extension data	Extension with PlGF
AUROC (95% CI)	0.88 (0.84–0.92)	0.82 (0.76–0.87)	0.81 (0.75–0.86)	0.73 (0.60–0.86)	0.77 (0.72–0.82)	0.67 (0.58-0.76)	0.67 (0.59-0.75)
Calibration slope	1	0.70 (0.63-0.76)	0.69 (0.63-0.75)	0.68 (0.56-0.79)	0.67 (0.59-0.74)	-	-
LR at risk cut-off (≥ 30%)		18.07 [11.60-28·16]	17.03 [10.97-26.43]	23.4 [14.83-36.79]	5·9 [4·23-8·35]		
NRI							0.02
IDI							-0.06

Bibliography

1. Magee LA, Pels A, Helewa M, Rey E, von Dadelszen P. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy. *Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health*. 2014;4(2):105-145.

2. von Dadelszen P, Magee LA. Pre-eclampsia: An update. *Curr Hypertens Rep*. 2014;16(8):454.

3. Hutcheon JA, Lisonkova S, Joseph KS. Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2011;25(4):391-403. doi: 10.1016/j.bpobgyn.2011.01.006.

4. Gillon TER, Pels A, von Dadelszen P, MacDonell K, Magee LA. Hypertensive disorders of pregnancy: A systematic review of international clinical practice guidelines. *PloS one*. 2014;9(12):e113715. doi: 10.1371/journal.pone.0113715.

5. Tranquilli AL, Dekker G, Magee L, et al. The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. *Pregnancy hypertension*. 2014;4(2):97-104. doi: 10.1016/j.preghy.2014.02.001.

6. World Health Organization, NCBI Bookshelf. *WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia.* Geneva, Switzerland: World Health Organization; 2011.

7. Task Force on Hypertension in Pregnancy, American College of Obstetricians and Gynecologists. Hypertension in pregnancy. report of the american college of obstetricians and gynecologists' task force on hypertension in pregnancy. *Obstet Gynecol*. 2013;122(5):1122.

8. Bennett S, Bourgeois C. HDP CPG working group, association of ontario midwives. hypertensive disorders of pregnancy. 2012;Clinical Practice Guideline no. 15.

9. National Institute for Health and Clinical Excellence (Great Britain), NCBI Bookshelf, Royal College of Midwives (Great Britain), National Collaborating Centre for Women's and Children's Health (Great Britain). *Hypertension in pregnancy: The management of hypertensive disorders during pregnancy.* Vol no. 107; no. 107. London: Royal College of Obstetricians and Gynaecologists; 2011.

10. Khan KS, Wojdyla D, Say L, Gülmezoglu AM, Van Look PF. WHO analysis of causes of maternal death: A systematic review. *The Lancet*. 2006;367(9516):1066-1074. doi: 10.1016/S0140-6736(06)68397-9.

11. von Dadelszen P, Firoz T, Donnay F, et al. Preeclampsia in low and middle income countries-health services lessons learned from the PRE-EMPT (PRE-eclampsia-eclampsia monitoring, prevention and treatment) project. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstétrique et gynécologie du Canada : JOGC*. 2012;34(10):917-926.

12. Wulf S, Johns N, Lozano R. Non-fatal burden of maternal conditions: Country-level results from the GBD 2010 study. *Lancet*. 2013;381(S2):149-149. doi: 10.1016/S0140-6736(13)61403-8.

13. Steegers EA, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. *The Lancet*. 2010;376(9741):631-644.

14. Redman CWG, Sargent IL. Placental stress and pre-eclampsia: A revised view. *Placenta*. 2009;30:38-42. doi: 10.1016/j.placenta.2008.11.021.

15. Ukah U, Payne B, Hoodbhoy Z, von Dadelszen P. Risk factors and predictors of pre-eclampsia. In: Magee L, von Dadelszen P, Stones W, Mathai M, eds. *The FIGO textbook of pregnancy hypertension.* First ed. London: The Global Library of Women's Medicine (GLOWM); 2016:75-100.

16. Webster LM, Conti-Ramsden F, Seed PT, Webb AJ, Nelson-Piercy C, Chappell LC. Impact of Antihypertensive Treatment on Maternal and Perinatal Outcomes in Pregnancy Complicated by Chronic Hypertension: A Systematic Review and Meta-Analysis. *J Am Heart Assoc.* 2017;6(5). doi: 10.1161/JAHA.117.005526.

17. Magee LA, von Dadelszen P, Rey E, et al. Less-tight versus tight control of hypertension in pregnancy. *N Engl J Med*. 2015;372(5):407-417.

18. Abalos E, Duley L, Steyn DW. Antihypertensive drug therapy for mild to moderate hypertension during pregnancy. *The Cochrane database of systematic reviews*. 2014(2):CD002252.

19. von Dadelszen P, Menzies JM, Payne B, Magee LA, PIERS (Pre-eclampsia Integrated Estimate of RiSk) Study Group. Predicting adverse outcomes in women with severe pre-eclampsia. *Semin Perinatol*. 2009;33(3):152-157. doi: 10.1053/j.semperi.2009.02.009.

20. Broekhuijsen K, van Baaren G.J, van Pampus M.G, Ganzevoort W, 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*. 2015; 385: 2492-2501.

21. Steyerberg EW, SpringerLink ebooks - Mathematics and Statistics. *Clinical prediction models: A practical approach to development, validation, and updating.* 1. Aufl. ed. New York, NY: Springer; 2009. 10.1007/978-0-387-77244-8.

22. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ: British Medical Journal*. 2009;338(7707):1373-1377.

23. Altman DG, Vergouwe Y, Royston P, Karel G M Moons. Prognosis and prognostic research: Validating a prognostic model. *BMJ: British Medical Journal*. 2009;338(7708):1432-1435.

24. von Dadelszen P, Payne B, Li J, et al. Prediction of adverse maternal outcomes in pre-eclampsia: Development and validation of the fullPIERS model. *The Lancet*. 2011;377(9761):219-227. doi: 10.1016/S0140-6736(10)61351-7.

25. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. external validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698.

26. Menzies J, Magee LA, MacNab YC, et al. Current CHS and NHBPEP criteria for severe preeclampsia do not uniformly predict adverse maternal or perinatal outcomes. *Hypertension in Pregnancy*. 2007;26(4):447-462. doi: 10.1080/10641950701521742.

27. Thangaratinam S, Coomarasamy A, O'Mahony F, et al. Estimation of proteinuria as a predictor of complications of pre-eclampsia: A systematic review. *BMC medicine*. 2009;7(1):10-10. doi: 10.1186/1741-7015-7-10.

28. Thangaratinam S, Koopmans CM, Iyengar S, et al. Accuracy of liver function tests for predicting adverse maternal and fetal outcomes in women with preeclampsia: A systematic review. *Acta Obstet Gynecol Scand*. 2011;90(6):574-585.

29. Koopmans CM, van Pampus MG, Groen H, Aarnoudse JG, van den Berg PP, Mol BWJ. Accuracy of serum uric acid as a predictive test for maternal complications in pre-eclampsia: Bivariate meta-analysis and decision analysis. *European Journal of Obstetrics and Gynecology*. 2009;146(1):8-14. doi: 10.1016/j.ejogrb.2009.05.014.

30. von Schmidt auf Altenstadt, Joost F., Hukkelhoven CWPM, van Roosmalen J, Bloemenkamp KWM. Pre-eclampsia increases the risk of postpartum haemorrhage: A nationwide cohort study in the netherlands. *PloS one*. 2013;8(12):e81959.

31. Honest H, Khan KS. Reporting of measures of accuracy in systematic reviews of diagnostic literature. *BMC health services research*. 2002;2(1):4-4. doi: 10.1186/1472-6963-2-4.

32. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280. doi: 10.7326/0003-4819-158-4-201302190-00009.

33. Al-Rubaie Z, Askie L, Ray J, Hudson H, Lord S. 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: An International Journal of Obstetrics & Gynaecology*. 2016;123(9):1441-1452. doi: 10.1111/1471-0528.14029.

34. Deeks JJ, Altman DG. Diagnostic tests 4: Likelihood ratios. *BMJ*. 2004;329(7458):168-169. doi: 10.1136/bmj.329.7458.168.

35. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36. doi: 10.1148/radiology.143.1.7063747.

36. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MGM, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making*. 2008;28(5):621-638. doi: 10.1177/0272989X08319957.

37. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. *Medical Decision Making*. 1993;13(4):313-321. doi: 10.1177/0272989X9301300408.

38. National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy. Report of the national high blood pressure education program working group on high blood pressure in pregnancy. *Obstet Gynecol*. 2000;183(1):s1-s22. doi: 10.1067/mob.2000.107928.

39. Ben Salem F, Ben Salem K, Grati L, et al. Risk factors for eclampsia: A case-control study. *Annales françaises d'anesthèsie et de rèanimation*. 2003;22(10):865.

40. Ankumah N, Cantu J, Jauk V, et al. Risk of adverse pregnancy outcomes in women with mild chronic hypertension before 20 weeks of gestation. *Obstet Gynecol*. 2014;123(5):966-972. doi: 10.1097/AOG.0000000000000205.

41. Witlin AG, Saade GR, Mattar F, Sibai BM. Risk factors for abruptio placentae and eclampsia: Analysis of 445 consecutively managed women with severe preeclampsia and eclampsia. *Obstet Gynecol*. 1999;180(6):1322-1329. doi: 10.1016/S0002-9378(99)70014-1.

42. Payne B, Magee LA, Magee LA, et al. PIERS proteinuria: Relationship with adverse maternal and perinatal outcome. *Journal of Obstetrics and Gynaecology Canada*. 2011;33(6):588-597.

43. Bouzari Z, Javadiankutenai M, Darzi A, Barat S. Does proteinura in preeclampsia have enough value to predict pregnancy outcome? *Clin Exp Obstet Gynecol*. 2014;41(2):163.

44. Gangaram R, Naicker M, Moodley J. Comparison of pregnancy outcomes in women with hypertensive disorders of pregnancy using 24-hour urinary protein and urinary microalbumin to creatinine ratio. *International Journal of Gynecology and Obstetrics*. 2009;107(1):19-22. doi: 10.1016/j.ijgo.2009.05.023.

45. Hall DR, Odendaal HJ, Steyn DW, Grové D. Urinary protein excretion and expectant management of early onset, severe pre-eclampsia. *International Journal of Gynecology and Obstetrics*. 2002;77(1):1-6. doi: 10.1016/S0020-7292(02)00008-5.

46. Schiff E, Friedman SA, Kao L, Sibai BM. The importance of urinary protein excretion during conservative management of severe preeclampsia. *Obstet Gynecol*. 1996;175(5):1313-1316. doi: 10.1016/S0002-9378(96)70047-9.

47. Morris RK, Riley RD, Doug M, Deeks JJ, Kilby MD. Diagnostic accuracy of spot urinary protein and albumin to creatinine ratios for detection of significant proteinuria or adverse pregnancy outcome in patients with suspected pre-eclampsia: Systematic review and meta-analysis. *BMJ: British Medical Journal*. 2012;345(7866):14-14. doi: 10.1136/bmj.e4342.

48. Yen T, Payne B, Qu Z, et al. Using clinical symptoms to predict adverse maternal and perinatal outcomes in women with preeclampsia: Data from the PIERS (pre-eclampsia integrated estimate of RiSk) study. *Journal of Obstetrics and Gynaecology Canada*. 2011;33(8):803-809. doi: 10.1016/S1701-2163(16)34983-0.

49. Aziz N, Kumar S. Clinical symptoms and laboratory parameters do not predict adverse maternal and fetal outcomes in HELLP. *Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health*. 2011;1(2):132-136. doi: 10.1016/j.preghy.2011.01.008.

50. Millman AL, Payne B, Qu Z, et al. Oxygen saturation as a predictor of adverse maternal outcomes in women with preeclampsia. *Journal of Obstetrics and Gynaecology Canada*. 2011;33(7):705-714. doi: 10.1016/S1701-2163(16)34955-6.

51. Thangaratinam S, Gallos ID, Meah N, et al. How accurate are maternal symptoms in predicting impending complications in women with preeclampsia? A systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2011;90(6):564.

52. Laskin S, Payne B, Hutcheon JA, et al. The role of platelet counts in the assessment of inpatient women with preeclampsia. *Journal of Obstetrics and Gynaecology Canada*. 2011;33(9):900-908.

53. Yucesoy G, Cakiroglu Y, Bodur H, Ozkan S, Tan T. An analysis of HELLP syndrome cases: Does platelet count predict adverse maternal and fetal outcomes in women with HELLP syndrome? *Arch Gynecol Obstet*. 2011;283(5):941-945. doi: 10.1007/s00404-010-1480-7.

54. Thangaratinam, Ismail K, Sharp, Coomarasamy, Khan, Tests in Prediction of Pre-eclampsia Severity review group. Accuracy of serum uric acid in predicting complications of pre-eclampsia: A systematic review. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2006;113(4):369-378.

55. Kozic JR, Benton SJ, Hutcheon JA, et al. Abnormal liver function tests as predictors of adverse maternal outcomes in women with preeclampsia. *Journal of Obstetrics and Gynaecology Canada*. 2011;33(10):995-1004.

56. Scazzochio E, Munmany M, Garcia L, et al. Prognostic role of maternal neutrophil gelatinase-associated lipocalin in women with severe early-onset preeclampsia. *Fetal Diagn Ther*. 2014;35(2):127.

57. Ghosh SK, Raheja S, Tuli A, Raghunandan C, Agarwal S. Association between placental growth factor levels in early onset preeclampsia with the occurrence of postpartum hemorrhage: A prospective cohort study. *Pregnancy hypertension*. 2012;2(2):115-122. doi: 10.1016/j.preghy.2011.11.006.

58. Leaños-Miranda A, Campos-Galicia I, Ramírez-Valenzuela KL, Chinolla-Arellano ZL, Isordia-Salas I. Circulating angiogenic factors and urinary prolactin as predictors of adverse outcomes in women with preeclampsia. *Hypertension*. 2013;61(5):1118-1125.

59. Palomaki GE, Haddow JE, Haddow HRM, et al. Modeling risk for severe adverse outcomes using angiogenic factor measurements in women with suspected preterm preeclampsia. *Prenat Diagn*. 2015;35(4):386-393. doi: 10.1002/pd.4554.

60. Saleh L, Verdonk K, Jan Danser AH, et al. The sFlt-1/PlGF ratio associates with prolongation and adverse outcome of pregnancy in women with (suspected) preeclampsia: Analysis of a high-risk cohort. *European Journal of Obstetrics and Gynecology*. 2016;199:121-126. doi: 10.1016/j.ejogrb.2016.02.013.

61. Rana S, Powe CE, Salahuddin S, et al. Angiogenic factors and the risk of adverse outcomes in women with suspected preeclampsia. *Circulation*. 2012;125(7):911-919.

62. Payne BA, Hutcheon JA, Ansermino JM, et al. A risk prediction model for the assessment and triage of women with hypertensive disorders of pregnancy in low-resourced settings: The miniPIERS (pre-eclampsia integrated estimate of RiSk) multi-country prospective cohort study. *PLoS Medicine*. 2014;11.

63. Payne BA, Hutcheon JA, Dunsmuir D, et al. Assessing the incremental value of blood oxygen saturation (SpO(2)) in the miniPIERS (pre-eclampsia integrated estimate of RiSk) risk prediction model. *Journal of obstetrics and gynaecology Canada : JOGC = Journal*

d'obstétrique et gynécologie du Canada : JOGC. 2015;37(1):16-24. doi: 10.1016/S1701-2163(15)30358-3.

64. Chan P, Brown M, Simpson JM, Davis G. Proteinuria in pre-eclampsia: How much matters? *BJOG: An International Journal of Obstetrics & Gynaecology*. 2005;112(3):280-285. doi: 10.1111/j.1471-0528.2004.00395.x.

65. Girling JC, Dow E, Smith JH. Liver function tests in pre-eclampsia: Importance of comparison with a reference range derived for normal pregnancy. *BJOG: An International Journal of Obstetrics and Gynaecology*. 1997;104(2):246-250. doi: 10.1111/j.1471-0528.1997.tb11054.x.

66. Allotey J, Marlin N, Mol BW, et al. Development and validation of prediction models for risk of adverse outcomes in women with early-onset pre-eclampsia: Protocol of the prospective cohort PREP study. *Diagnostic and Prognostic Research*. 2017;1(1). doi: 10.1186/s41512-016-0004-8.

67. Redman CW, Sargent IL. Latest advances in understanding preeclampsia. *Science*. 2005;308(5728):1592-1594.

68. Vaisbuch E, Whitty JE, Hassan SS, et al. Circulating angiogenic and antiangiogenic factors in women with eclampsia. *Obstet Gynecol*. 2011;204(2):152.e1-152.e9. doi: 10.1016/j.ajog.2010.08.049.

69. Akolekar R, Zaragoza E, Poon LCY, Pepes S, Nicolaides KH. Maternal serum placental growth factor at 11 + 0 to 13 + 6 weeks of gestation in the prediction of pre-eclampsia. *Ultrasound in Obstetrics and Gynecology*. 2008;32(6):732-739.

70. Sibiude J, Guibourdenche J, Dionne M, et al. Placental growth factor for the prediction of adverse outcomes in patients with suspected preeclampsia or intrauterine growth restriction. *PloS one*. 2012;7(11):e50208.

71. Alahakoon TI, Zhang W, Trudinger BJ, Lee VW. Discordant clinical presentations of preeclampsia and intrauterine fetal growth restriction with similar pro- and anti-angiogenic profiles. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2014;27(18):1854-1859. doi: 10.3109/14767058.2014.880882.

72. Baltajian K, Bajracharya S, Salahuddin S, et al. Sequential plasma angiogenic factors levels in women with suspected preeclampsia. *Obstet Gynecol*. 2016;215(1):89.e1-89.e10. doi: 10.1016/j.ajog.2016.01.168.

73. Álvarez-Fernández I, Prieto B, Rodríguez V, Ruano Y, Escudero AI, Álvarez FV. N-terminal pro B-type natriuretic peptide and angiogenic biomarkers in the prognosis of adverse outcomes in women with suspected preeclampsia. *Clinica Chimica Acta*. 2016;463:150-157. doi: 10.1016/j.cca.2016.10.033.

74. Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: Available, but far from applicable. *Obstet Gynecol*. 2016;214(1):79-90.e36. doi: 10.1016/j.ajog.2015.06.013.

75. Frampton GK, Jones J, Rose M, Payne L. Placental growth factor (alone or in combination with soluble fms-like tyrosine kinase 1) as an aid to the assessment of women with suspected pre-eclampsia: Systematic review and economic analysis. *Health Technol Assess*. 2016;20(87):1-160. doi: 10.3310/hta20870.

76. Ukah UV, Mbofana F, Rocha BM, et al. Diagnostic performance of placental growth factor in women with suspected preeclampsia attending antenatal facilities in maputo, mozambique. *Hypertension*. 2017;69(3):469-474. doi: 10.1161/HYPERTENSIONAHA.116.08547.

77. Molvarec A, Gullai N, Stenczer B, Fügedi G, Nagy B, Rigó J, János. Comparison of placental growth factor and fetal flow doppler ultrasonography to identify fetal adverse outcomes in women with hypertensive disorders of pregnancy: An observational study. *BMC pregnancy and childbirth*. 2013;13(1):161-161.

78. De Oliveira L, Peraçoli JC, Peraçoli MT, et al. sFlt-1/PlGF ratio as a prognostic marker of adverse outcomes in women with early-onset preeclampsia. *Pregnancy hypertension*. 2013;3(3):191.

79. Gómez-Arriaga PI, Herraiz I, López-Jiménez EA, Escribano D, Denk B, Galindo A. Uterine artery doppler and sFlt-1/PlGF ratio: Prognostic value in early-onset pre-eclampsia: Prognostic value of sFlt-1/PlGF and UtA doppler. *Ultrasound in Obstetrics & Gynecology*. 2014;43(5):525-532. doi: 10.1002/uog.13224.

80. Rana S, Hacker MR, Modest AM, et al. Circulating angiogenic factors and risk of adverse maternal and perinatal outcomes in twin pregnancies with suspected preeclampsia. *Hypertension*. 2012;60(2):451-458.

81. Chaiworapongsa T, Romero R, Savasan ZA, et al. Maternal plasma concentrations of angiogenic/anti-angiogenic factors are of prognostic value in patients presenting to the obstetrical triage area with the suspicion of preeclampsia. *Journal of Maternal-Fetal and Neonatal Medicine*. 2011;24(10):1187-1207. doi: 10.3109/14767058.2011.589932.

82. Chaiworapongsa T, Romero R, Korzeniewski SJ, et al. Plasma concentrations of angiogenic/anti-angiogenic factors have prognostic value in women presenting with suspected preeclampsia to the obstetrical triage area: A prospective study. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2014;27(2):132-144.

83. Chappell LC, Duckworth S, Seed PT, et al. Diagnostic accuracy of placental growth factor in women with suspected preeclampsia: A prospective multicenter study. *Circulation*. 2013;128(19):2121-2131.

84. Moore Simas T. Angiogenic factors for the prediction of preeclampsia in high-risk women. *Obstet Gynecol*. 2007;197(3):244.e1-244.e8. doi: 10.1016/j.ajog.2007.06.030.

85. Salahuddin S, Wenger JB, Zhang D, Thadhani R, Karumanchi SA, Rana S. KRYPTOR-automated angiogenic factor assays and risk of preeclampsia-related adverse outcomes. *Hypertension in pregnancy*. 2016;35(3):330-345. doi: 10.3109/10641955.2016.1148162.

86. Woelkers DA, von Dadelszen P, Sibai B. 482: Diagnostic and prognostic performance of placenta growth factor (PLGF) in women with signs or symptoms of early preterm preeclampsia. *Obstet Gynecol*. 2016;214(1):S264. doi: 10.1016/j.ajog.2015.10.525.

87. Rana S, Schnettler WT, Powe C, et al. Clinical characterization and outcomes of preeclampsia with normal angiogenic profile. *Hypertension in Pregnancy*. 2013;32(2):189-201.

88. Jääskeläinen T, Heinonen S, Kajantie E, et al. Cohort profile: The finnish genetics of pre-eclampsia consortium (FINNPEC). *BMJ Open*. 2016;6(11):e013148.

89. Kaartokallio T, Klemetti MM, Timonen A, et al. Microsatellite polymorphism in the heme oxygenase-1 promoter is associated with nonsevere and late-onset preeclampsia. *Hypertension*. 2014;64(1):172-177. doi: 10.1161/HYPERTENSIONAHA.114.03337.

90. Ganzevoort W, Rep A, de Vries JIP, Bonsel GJ, Wolf H, PETRA-investigators. Prediction of maternal complications and adverse infant outcome at admission for temporizing management of early-onset severe hypertensive disorders of pregnancy. *Obstet Gynecol*. 2006;195(2):495-503. doi: 10.1016/j.ajog.2006.02.012.

91. Thangaratinam S, Allotey J, Marlin N, et al. Prediction of complications in early-onset pre-eclampsia (PREP): Development and external multinational validation of prognostic models. *BMC MEDICINE*. 2017;15.

92. Cummings P. Missing data and multiple imputation. *JAMA Pediatrics*. 2013;167(7):656-661. doi: 10.1001/jamapediatrics.2013.1329.

93. Rubin DB. *Multiple imputation for nonresponse in surveys*. 99; 99th ed. New York: Wiley; 1987; 2009; 2002.

94. Buuren v, Stef, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011;45(3).

95. Rippe RCA, den Heijer M, le Cessie S. Imputation of missing data. *Ned Tijdschr Geneeskd*. 2013;157(18):A5539.

96. Held U, Kessels A, Garcia Aymerich J, et al. Methods for handling missing variables in risk prediction models. *Am J Epidemiol*. 2016;184(7):545-551. doi: 10.1093/aje/kwv346.

97. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi: 10.1097/EDE.0b013e3181c30fb2.

98. Agrawal S, Maitra N. Prediction of adverse maternal outcomes in preeclampsia using a risk prediction model. *J Obstet Gynaecol India*. 2016;66(Suppl 1):104.

99. Ukah UV, Payne B, Lee T, Magee LA, von Dadelszen P. 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.

100. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med*. 2008;149(10):751.

101. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design. *J Natl Cancer Inst*. 2008;100(20):1432-1438. doi: 10.1093/jnci/djn326.

102. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5):475-483. doi: 10.1016/j.jclinepi.2004.06.017.

103. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi: 10.1016/j.jclinepi.2014.06.018.

104. Vergouwe Y, Moons C, Steyerberg E. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971-980. doi: 10.1093/aje/kwq223.

105. Jonathan A C Sterne, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ: British Medical Journal*. 2009;339(7713):157-160.

106. Raghunathan TE. What do we do with missing data? some options for analysis of incomplete data. *Annu Rev Public Health*. 2004;25(1):99-117.

107. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087-1091.

108. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. *J Clin Epidemiol*. 2008;61(11):1085-1094.

109. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.

110. Smith GCS, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. *Am J Epidemiol*. 2014;180(3):318-324.

111. Harrell J,F E., Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361.

112. Karel G M Moons, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ: British Medical Journal*. 2009;338(7709):1487-1490.

113. Van Calster B, Van Hoorde K, Vergouwe Y, et al. Validation and updating of risk models based on multinomial logistic regression. *Diagnostic and Prognostic Research*. 2017;1(1).

114. Pastor-Barriuso R, Ascunce N, Ederra M, et al. Recalibration of the gail model for predicting invasive breast cancer risk in spanish women: A population-based cohort study. *Breast Cancer Res Treat*. 2013;138(1):249-259.

115. Schuetz P, Koller M, Christ-Crain M, et al. Predicting mortality with pneumonia severity scores: Importance of model recalibration to local settings. *Epidemiol Infect*. 2008;136(12):1628-1637.

116. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the united kingdom. *Crit Care Med*. 2006;34(5):1378-1388.

117. Saffer C, Olson G, Boggess KA, et al. Determination of placental growth factor (PlGF) levels in healthy pregnant women without signs or symptoms of preeclampsia. *Pregnancy hypertension*. 2013;3(2):124-132.

Appendices

Appendix A List of validation data studies, collaborators and other contributors

Studies	Collaborators
fullPIERS working group	Peter von Dadelszen(Principal Investigator - PI), Beth Payne, Jing Li, J Mark Ansermino, Fiona Broughton Pipkin, Anne-Marie Côté, M Joanne Douglas, Andrée Gruslin, Jennifer A Hutcheon, K S Joseph, Phillipa M Kyle, Tang Lee, Pamela Loughna, Jennifer M Menzies, Mario Merialdi, Alexandra L Millman, M Peter Moore, Jean-Marie Moutquin, Annie B Ouellet, Graeme N Smith, James J Walker, Keith R Walley, Barry N Walters, Mariana Widmer, Shoo K Lee, James A Russell, Laura A Magee, for the PIERS Study Group
BCW	Peter von Dadelszen (PI) Ryan Ingram, Vivien Cao, Lilian Cao, Phae Azurin, Katherine Thomas, Matthew Haslam, Elizabeth Wilcox, Benita Okocha, Emma von Dadelszen
FINNPEC	Hannelle Laivouri (PI) Hanna K. Karjalainen, Kortelainen, Eija M Kortelainen
PELICAN	Lucy Chappell (PI) Paul T Seed, Frances Inez Conti-Ramsden
Alere PETRA	Paul Sheard (PI) Doug Woelkers, Kenneth Kupfer, Baha Sibai
JRH, Oxford	Manu Vatish (PI) Christopher Redman, Layla Lavallee, Fiona Goddard, Pawel Szafranski, Karen Melham
PREP	Shakila Thangaratinam (PI) John Allotey, Julie Dodds
miniPIERS	Beth A. Payne, Jennifer A. Hutcheon, J. Mark Ansermino, David R. Hall, Zulfiqar A. Bhutta, Shereen Z. Bhutta, Christine Biryabarema, William A. Grobman, Henk Groen, Farizah Haniff, Jing Li, Laura A. Magee, Mario Merialdi, Annettee Nakimuli, Ziguang Qu, Rozina Sikandar, Nelson Sass, Diane Sawchuck, D. Wilhelm Steyn, Mariana Widmer, Jian Zhou, Peter von Dadelszen, for the miniPIERS Study Working Group
Dutch PETRA	J.Wessel Ganzevoort (PI)

HYPITAT	Henk Groen (PI) Eva Zwerbroek, Maureen Franssen, Kim Broekhuijsen
VIPER – PRE-EMPT	Beth Payne, Sharla Drebit, Tang Lee, Larry Li, Dane De Silva, Marianne Vidler, Asif Raza, Sumedha Sharma, Jeffrey Bone, Chirag Kariya, Alison Dube, Domena Tu, Dustin Dunsmuir, KS Joseph, Sarka Lisonkova, Joanne Kirton, Anna Hutfield, Helen Ryan, Yuxiang, Michelle La, Kristina Arion, Maggie Kinshella, Kevin Au, Mansun, Tashya De Silva, Meera Madhavan, Tabassum Firoz, Tatenda P. Makanga
CHIPS	Bill Hague, Suzette Coat, Nestor Demianczuk, Venu Jain, Carmen Young, Cheryl Lux, Cora Fanning, Jean-Marie Moutquin, Anne-Marie Cote, Jany Rodrigue, Michelle Hladunewich, Anna Rogowsky, Anny Gonzalez, Susan Jackson, Carly Minuk, Keith Still, Chris Hotz, Evelyne Rey, Heather Clark, Patty Waddell, Eileen O'shee, Ruth Rennicks White, Nathalie Rybak, Joel G Ray, Leanne De Souza, Jenita Chrysostoum, Fred Kirss, Kristina Rull, Ruth Hughes, Di Leishman, Wessel Ganzevoort, Eline van den Akker, Steven Koenen, Steve A Walkinshaw, Mark Clement-Jones, Julie Wray, Rachael OKeeffe, Michelle dower, James Thornton, Ann Smith, Maria Yvette davis, Hannah Driver, Derek Tuffnell, Diane Farrar, Jennifer Syson, Declan Ryan-Wakeling, David Churchill, Julie Icke, Kate Cheshire, Katherine Cheshire. Amelia Turner, Laura Gardiner, Helen M. Cameron, Eileen Walton, Mathew Princy, Pauline Atkinson, Lynne Palmer, Catherine Nelson-Piercy, Hayley Tarft, Annette Briley, Meena Khandelwal, Gunda Simkins
Clinicians (informal discussions)	Dr. Nancy Kent, Dr. Amanada Skoll, Dr. Wee-Shian Chan

Appendix B PIERS maternal adverse outcomes

Outcome	Definition
Maternal	
Mortality	Maternal death occurring within six weeks of pregnancy or if later, attributable to complications of pre-eclampsia
Hepatic dysfunction	INR >1.2 in the absence if DIC or treatment of Warfarin (DIC is defined as having both: abnormal bleeding and consumptive coagulopathy (i.e., low platelets, abnormal peripheral blood film, or one or more of the following: increased INR, increased PTT, low fibrinogen, of increased fibrin degradation products that are outside normal non-pregnancy ranges))
Hepatic hematoma or rupture	Blood collection under the hepatic capsule as confirmed by ultrasound or laparotomy
Glasgow coma score < 13	Based on GCS scoring system: Teasdale G, Jennet B. Assessment of coma and impaired consciousness: a practical scale. *Lancet* 1974; **2**:81-83

Stroke	Acute neurological event with deficits lasting longer than 48 hours
Cortical Blindness	Loss of visual acuity in the presence of intact papillary response to light
Reversible Ischaemic Neurologic Deficit (RIND)	Cerebral ischaemia lasting longer than 24 hrs but less than 48 hours revealed through clinical examination
Retinal detachment	Separation of the inner layers of the retina from the underlying retinal pigment epithelium (RPE, choroid) and is diagnosed by ophthalmological exam
Acute renal insufficiency	For women with an underlying history of renal disease: defined as creatinine >200 µM; for patients with no underlying renal disease: defined as creatinine >150 µM
Dialysis	Including haemodialysis and peritoneal dialysis
Platelet count < 50,000 without blood transfusion (thrombocytopoenia)	Measurement of platelet count recorded as less than 50,000 without patient being given a blood transfusion
Transfusion of blood products	Includes transfusion of any units of blood products: fresh frozen plasma (FFP), platelets,

	red blood cells (RBCs), cryoprecipitate (cryo) or whole blood
Positive inotropic support	The use of vasopressors to maintain a sBP > 90 mmHg or Mean Arterial pressure > 70 mmHg
Myocardial ischaemia/infarction	ECG changes (ST segment elevation or depression) without enzyme changes AND/OR any one of the following: 1)Development of new pathologic Q waves on serial ECGs. The patient may or may not remember previous symptoms. Biochemical markers of myocardial necrosis may have normalized, depending on the length of time that has passed since the infarct developed. 2) Pathological findings of an acute, healed or healing MI 3) Typical rise and gradual fall (troponin) or more rapid rise and fall (CK-MB) of biochemical markers of myocardial necrosis with at least one of the following: a) ischaemic symptoms; b) development of pathologic Q waves on the ECG; c) ECG changes indicative of ischaemia (ST segment

	elevation or depression); or d) coronary artery intervention (e.g., coronary angioplasty)
Require >50% oxygen for greater than one hour	Oxygen given at greater than 50% concentration based on local criteria for longer than 1 hour
Intubation other than for Cesarean section	Intubation may be by ventilation, EIT or CPAP
Pulmonary Oedema	Clinical diagnosis with x-ray confirmation or requirement of diuretic treatment and SaO_2 <95%
Eclampsia (Eclamptic seizures)	The occurrence of generalised convulsions during pregnancy, labour, or within 7 days of delivery in the absence of epilepsy or another condition predisposing to convulsions. A seizure or convulsion can be a sudden, violent, uncontrollable contraction of a group of muscles. A seizure can also be subtler, consisting of only a brief loss of contact.
Other adverse events	Placental abruption Severe ascites Bell's palsy Hysterectomy

Appendix C Search strategy for review of maternal outcomes prediction from HDPs

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present> Search Strategy:

---

1 exp Hypertension, Pregnancy-Induced/ (31573)

2 (HDP or HDPs).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (940)

3 (preeclamp* or pre-eclamp*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (35884)

4 ((Chronic hypertens* or essential hypertens* or preexisting hypertens* or pre-existing hypertens*) adj3 (pregnan* or gestation*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (653)

5 HELLP Syndrome/ (1609)

6 HELLP.mp. (2473)

7 or/1-6 (41093)

8 nomograms/ (1952)

9 Models, statistical/ (78055)

10 logistic models/ (108873)

11 "Predictive Value of Tests"/ (168044)

12 Risk assessment/ (207143)

13 clinical risk assessment*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (302)

14 prognos* model*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2879)

15 predict* model*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (20780)

16 (AUC or AUROC or area under the receiver or ROC or ROCs or Receiver operating curve*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (96577)

17 sensitivit*.mp. (935762)

18 specificit*.mp. (926124)

19 (LR* or likelihood ratio*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (45658)

20 negative predictive value*.mp. (34619)

21 positive predictive value*.mp. (35114)

22 or/8-21 (1990826)

23 ((risk* or predict* or prognos*) adj6 (adverse or complication* or outcome* or event* or situation*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (286836)

24 ((risk* or predict* or prognos*) adj6 (morbid* or mortality)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (104862)

25 ((risk* or predict* or prognos*) adj6 (Hepatic or GCS or Glasgow or Stroke or Cortical or RIND or retinal or Dialysis or renal or PIS or Positive inotropic support or Infusion or Myocardial or MI or Intubation or thrombocytop?nia)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (128207)

26 23 or 24 or 25 (466662)

27 7 and 22 and 26 (828)

Appendix D Search strategy for review on PlGF as a prognostic marker of HDPs

A. Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>
Search Strategy:

---

1 exp Hypertension, Pregnancy-Induced/ (33898)
2 (HDP or HDPs).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (1059)
3 (preeclamp* or pre-eclamp*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (38748)
4 ((Chronic hypertens* or essential hypertens* or preexisting hypertens* or pre-existing hypertens*) adj3 (pregnan* or gestation*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (686)
5 HELLP Syndrome/ (1682)
6 HELLP.mp. (2610)
7 or/1-6 (44307)
8 nomograms/ (2183)
9 Models, statistical/ (89271)
10 logistic models/ (124380)
11 "Predictive Value of Tests"/ (182314)
12 Risk assessment/ (228067)
13 clinical risk assessment*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (337)
14 prognos* model*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3291)
15 predict* model*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (23811)
16 (AUC or AUROC or area under the receiver or ROC or ROCs or Receiver operating curve*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (106924)
17 sensitivit*.mp. (1022881)
18 specificit*.mp. (1017136)
19 (LR* or likelihood ratio*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (52107)
20 negative predictive value*.mp. (37119)

21 positive predictive value*.mp. (37906)
22 or/8-21 (2197479)
23 ((risk* or predict* or prognos*) adj6 (adverse or complication* or outcome* or event* or situation*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (316116)
24 ((risk* or predict* or prognos*) adj6 (morbid* or mortality)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (116793)
25 ((risk* or predict* or prognos*) adj6 (Hepatic or GCS or Glasgow or Stroke or Cortical or RIND or retinal or Dialysis or renal or PIS or Positive inotropic support or Infusion or Myocardial or MI or Intubation or thrombocytop?nia)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (139161)
26 ((risk* or predict* or prognos*) adj6 (stillbirth or IUGR or intrauterine growth restriction or FGR or low birthweight or fetal growth restriction or delivery or preterm)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (21573)
27 or/23-26 (529782)
28 placental growth factor.mp. (1686)
29 plgf.mp. (1728)
30 Angiogenesis Inducing Agents/ (3504)
31 proangiogenic factor*.mp. (1025)
32 angiogenic factor*.mp. (8950)
33 or/28-32 (14054)
34 7 and 22 and 27 and 33 (82)

****