Counseling women with hypertensive disorders of pregnancy
van Oostwaard, M.F.

Citation for published version (APA):
van Oostwaard, M. F. (2015). Counseling women with hypertensive disorders of pregnancy

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

External validation of prognostic models for recurrence of hypertensive disorders of pregnancy in a pooled dataset of observational cohorts

M.F. van Oostwaard
E. Schuit
J. Langenveld
D.N.M. Papatsonis
M.A. Brown
R.N. Byaruhanga
S. Bhattacharya
D.M. Campbell
L.C. Chappell
F. Chiaffarino
I. Crippa
F. Facchinetti
S. Ferrazzani
E. Ferrazzi
E.A. Figueiró-Filho
I.P.M. Gaugler-Senden
C. Haavaldsen
J.A. Lykke
A.K. Mbah
V.M. Oliveira
L. Poston
C.W.G. Redman
R. Salim
L.J.M. Smits
B. Thilaganathan
P. Vergani
J. Zhang
B.W.J. Mol
W. Ganzvoort

Concept
Chapter 6

Abstract:

**Background:** Hypertensive disorders of pregnancy (HDP) complicate approximately 2 to 8 per cent of all pregnancies and are associated with potentially severe maternal and neonatal complications. Counseling on individual recurrence risk of HDP in a future pregnancy is important. Prognostic models for recurrence of HDP have been developed, but their generalizability and applicability is unclear due to a lack of externally validation studies.

**Objective:** We aimed to validate existing prognostic models for recurrence of hypertensive disorders of pregnancy (HDP) in a pooled cohort of observational studies.

**Study design.** Data from a previously performed individual participant data meta-analysis were used as validation population. Missing data were imputed using multiple imputation. We searched Medline for eligible prognostic models for recurrence of HDP. After applying the models in our validation population overall performance was assessed by calibration (calibration plots) and discrimination (c-statistic). If performance was poor, the existing models were updated by adjustment of intercept and/or overall slope.

**Results:** We used an individual patient database of 22 studies, containing data of 99,611 women with previous HDP. We validated 4 prognostic models for recurrence of HDP: a model for predicting pre-eclampsia (PE) and a model for predicting early onset PE (< 34 weeks) in women with previous PE and a delivery before 37 weeks or HELLP (N= 11,298), a model for predicting PE < 34 weeks’ gestation in women with previous PE or HELLP syndrome and a delivery before 34 weeks’ gestation (N= 3935), and a model predicting HDP in women with previous HDP and delivery between 34 and 37 weeks’ gestation (N= 5590). The c-statistics of the models were 0.61, 0.98, 0.57 and 0.58, respectively, indicating poor to very good discrimination. Calibration plots of the original models showed over- and underprediction at various risk ranges, which substantially improved after adjusting the intercept and slope for all models.

**Conclusion:** All prognostic models were able to perform accurate risk prediction after adjustment of the intercept and/or slope. However, discriminative ability of all models but one was limited. Updating the existing models with new predictors may improve discriminative ability of the prognostic models. Future research is needed to evaluate clinical applicability and implementation.
Introduction

Hypertensive disorders of pregnancy (HDP) complicate approximately 2 to 8 per cent of all pregnancies\(^1\)\(^-\)\(^3\) and comprise gestational hypertension (GH), preeclampsia (PE), superimposed PE and HELLP (Hemolysis, Elevated Liver enzymes and Low Platelets) syndrome, and in a varying percentage of cases they are related to intrauterine growth restriction. They are associated with potentially severe maternal and neonatal complications\(^1\)\(^-\)\(^3\). There is probably important heterogeneity in pathophysiology and clinical phenotype between subgroups and individual women. HDP can also have a major psychological impact on the woman and her family\(^4\). As such, counseling on individual recurrence risk of a hypertensive disease in a future pregnancy is important.

Many risk factors for recurrence of HDP have been identified, like parity, chronic hypertension and maternal age\(^5\). However, clinical practice and prognosis are multivariable. Therefore, performance of prediction based on single risk factors has been disappointing\(^5\). Information from multiple variables (predictors) originating from demographics, general history and clinical syndrome in the previous pregnancy, can be combined into a prognostic model. This way, prognostic models can help in the assessment of one's individual recurrence risk for HDP. A recent review shows that prognostic models are not implemented in clinical practice, despite the existence of many prognostic models, including models for recurrence of HDP\(^6\). One of the reasons is the lack of external validation of developed prognostic models. External validation is important for evaluation of generalizability and applicability in different populations\(^6\).

The use of individual participant data (IPD) combined from multiple studies is relatively new to prognostic research\(^7\). Combining data enlarges the study population and thereby increases the number of events per variable in a prognostic model, resulting in more reliable development or validation of prognostic model\(^8\). We have recently performed an IPD meta-analysis on recurrence of HDP\(^9\), which can be very useful in the external validation of existing prognostic models\(^10\). Therefore, the goal of this study is to validate existing prognostic models for recurrence of HDP using a pooled cohort of observational studies.
Materials and methods

Validation population
Our validation population consisted of individuals included in a recently published individual patient data meta-analysis. A detailed description of the literature search, study selection and data collection can be found elsewhere. Of one study, permission to use data was obtained after publication of the meta-analysis, adding data of 407 women with former PE and a delivery before 34 weeks to our database. Due to overlapping data of this study with the IPD of another study previously included, the latter study was excluded. Eventually, the individual patient data of 22 articles were included in this IPD meta-analysis, providing IPD of 99,611 women with a previous HDP. Study characters, risk of bias table and baseline characteristics are also shown in our prior article.

Prognostic models
A literature search was performed in the Medline electronic library (through PubMed). No language restrictions were applied. The search covered all records from inception until March 2015. The search strategy is based on the strategy used by Kleinrouweier et al. and is presented in Table 1. Reference lists of selected studies were checked to identify other studies of interest. All studies describing prognostic models for recurrence of HDP and pertaining before a subsequent pregnancy occurred, were considered eligible. Models developed to be applied during pregnancy were excluded, because interest was in predicting recurrence before a consecutive pregnancy.

Study selection shown in the flow chart in Figure 1. Our search identified 9 eligible pre-pregnant prognostic models for the recurrence of HDP. Three prognostic models were excluded because they contained more than one predictor not available in our IPD set. Two other models were excluded since the IPD had only very limited data on more than one of the required predictors. Four models were considered suitable for validation: 2 models by Seed et al., a model by Van Kuijk et al. and a model by Van Oostwaard et al. Characteristics of these 4 models, including regression formula, are presented in Table 2.
Table 1. Search strategy for existing prognostic models for HDP

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Validat*[tiab] OR Predict*[tiab] OR Rule*[tiab]</td>
</tr>
<tr>
<td>2.</td>
<td>Predict*[tiab] AND (Outcome*[tiab] OR Risk*[tiab] OR Model*[tiab])</td>
</tr>
<tr>
<td>4.</td>
<td>Decision*[tiab] AND (Model*[tiab] OR Clinical*[tiab] OR Logistic Model*[tiab])</td>
</tr>
<tr>
<td>6.</td>
<td>“risk score”[All fields] OR “prediction model”[All fields] OR “prediction rule”[All fields] OR “risk assessment”[All fields] OR “algorithm”[All fields]</td>
</tr>
<tr>
<td>7.</td>
<td># 1 OR #2 OR #3 OR #4 OR #5 OR #6</td>
</tr>
<tr>
<td>8.</td>
<td>preeclampsia*[tiab] OR pre-eclampsia*[tiab] OR pregnancy induced hypertension*[tiab] OR HELLP*[tiab]</td>
</tr>
<tr>
<td>9.</td>
<td>#7 AND #8</td>
</tr>
</tbody>
</table>

All four models were logistic regression models. The 2 models by Seed et al. predict PE at any gestational age (Seed1) and early onset PE (< 34 weeks' gestation, Seed2). Their development population included women with previous PE or HELLP, chronic hypertension, diabetes, antiphospholipid syndrome, chronic renal disease and in-pregnancy risk factors: multiple pregnancy, primigravidae with high Body Mass Index (BMI) (≥ 30 kg/m²) and abnormal uterine artery Doppler studies. The models were externally validated by the same investigators (in the same publication), which revealed a c-statistic of 0.66 for model Seed1 and 0.81 for model Seed2.

The model by Van Kuijk et al. predicts early onset PE (< 34 weeks' gestation) and was derived in a population of women with former PE or HELLP and a delivery before 34 weeks' gestation. External validation of the model by Van Kuijk revealed a c-statistic of 0.59 (95% CI: 0.45–0.73) and poor calibration. They recommended a weighted mean intercept of 0.59 for future use.

The model by Van Oostwaard et al. predicts HDP at any gestational age and was developed in a population with previous HDP in the late preterm period (34-37 weeks' gestation). The model was not externally validated. Internal validation revealed a c-statistic of 0.71 (95% CI 0.64–0.78).

Definitions

As primary outcome in the validation process we used the outcome definitions of the prognostic models. The outcome of the Seed1 model was: recurrence of preeclampsia, defined as gestational hypertension with proteinuria, according to the International
Society for the Study of Hypertension in Pregnancy (ISSHP). The outcome of the Seedz model was defined as: recurrent preeclampsia resulting in delivery before 34 weeks gestation. The outcome of the Van Kuijk model was defined as: recurrent early-onset preeclampsia, including the HELLP syndrome, according to the ISSHP criteria, resulting in delivery before 34 weeks of gestation. The outcome of the Van Oostwaard model was: recurrence of hypertensive disorders of pregnancy and included PE, superimposed PE, GH, HELLP syndrome and/or delivery of a SGA child. Preeclampsia was defined according to the ACOG criteria.

Figure 1. Inclusion of prognostic models
External validation of HDP recurrence models

**Statistical analysis**

Studies included in our IPD database that were missing more than 50% of the models’ predictors were excluded before performing the validation process. This resulted in the exclusion of 5 studies\(^{15,20,26,28,32}\) and data of the model of Van Kuijk\(^{11}\), which were excluded from the pooled dataset before validation of that particular model.

Not all predictors and outcomes had been collected in every study within the pooled cohort. Since imputation of missing values reduces bias, especially when data are selectively missing, it allows for the inclusion of all women in the analysis.\(^{42}\) We multiple imputed these missing values 10 times within each individual study. Within each study missing values were only imputed for variables for which information was not completely missing. Variables that were completely missing were initially not imputed. After combining all ten imputed datasets from all individual studies, we performed a single imputation in the total database to also impute those variables within each study that were completely missing (but were available in other studies included in the pooled cohort). This approach was chosen because the best discrimination and a calibration intercepts closest to the true value could be obtained by after imputation of predictor that was completely missing.\(^{43}\) Our pooled database had no information on use of folate (Seed’s model) or on fasting blood glucose (Van Kuijk model). Data on these predictors were obtained from the study groups that developed the models for the purpose of imputation.

For the validation of each of the identified prognostic models we created validation populations according to the gestational age and specific individual syndrome of the former pregnancy in the models’ development population (Table 2). For validation of the Seed models we selected women with previous PE and delivery before 37 weeks or women with previous HELLP syndrome (any GA) (11,298 women). Because the derivation population included women with other risk factors for HDP than previous HDP, we could not completely match populations. For the Van Kuijk model we selected women with previous PE or HELLP syndrome and a delivery before 34 weeks of gestation (3935 women). For the by Van Oostwaard model we selected women with previous HDP and delivery between 34 and 37 weeks gestational age (5590 women). Note that all statistical procedures, including selection of validation populations were performed after the multiple imputation procedure. Descriptive statistics were performed for each validation population separately. Data are presented as percentages or expressed as mean with standard deviation (SD) or median with interquartile range (IQR), as appropriate.
### Table 2. Characteristics of the 4 existing prognostic models

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Predictors</th>
<th>Outcome</th>
<th>The model</th>
<th>IPD data **</th>
</tr>
</thead>
</table>
| Seed 1<sup>st</sup> 2011 | Included in our IPD:  
• Former PE < 37 weeks  
• Former HELLP syndrome  
Not included in our IPD:  
• Chronic hypertension  
• Primigravidae & BMI ≥ 30 kg/m2  
• Type 1 or type 2 diabetes  
• Multiple pregnancy  
• Antiphospholipid syndrome  
• Chronic renal disease  
• Abnormal uterine artery Doppler studies | • Former PE / HELLP  
• Ethnicity  
• Chronic hypertension  
• Blood pressures at booking  
• Use of folate<sup>*</sup> | Recurrent PE any GA | Predicted Risk =  
\[
\frac{1}{1 + \exp\left(-1.2422 + 0.4061 \text{ (if CHT)} + 0.5071 \text{ (if DBP > 70)} - 0.3846 \text{ (if DBP > 90)} + 0.8890 \text{ (if SBP > 120)} + 0.7040 \text{ (if previous preeclampsia) - 0.4043 (if on folates) + 0.8311 (if mother is Indian, Bangladeshi, Pakistani, African, or Afro-Caribbean)}\right)}
\] | 11.298 |
| Seed 2<sup>nd</sup> 2011 | Same as above  
• Former PE / HELLP  
• Previous delivery < 34 wks  
• Ethnicity  
• Use of antihypertensive medication before subsequent pregnancy  
• Blood pressures at booking | • Former PE / HELLP  
• Previous delivery < 34 wks  
• Ethnicity  
• Blood pressures at booking | Recurrent PE < 34 wks | Predicted Risk =  
\[
\frac{1}{1 + \exp\left(-2.693 + 1.735 \text{ (if SBP > 140)} + 1.004 \text{ (if on antihypertensive therapy)} - 0.9790 \text{ (if previous preeclampsia in most recent pregnancy /HELLP/eclampsia)} + 2.121 \text{ (if previous preeclampsia with delivery before 34 weeks)} + 1.285 \text{ (if mother is Indian, Bangladeshi, Pakistani, African, or Afro-Caribbean)}\right)}
\] | 11.298 |
| Van Kuijk<sup>11</sup> 2011 | Former PE or HELLP  
And delivery < 34 wks | • Chronic hypertension  
• GA at former delivery  
• Former SGA  
• BMI  
• Fasting blood glucose<sup>*</sup> | Recurrent PE < 34 wks | Predicted Risk =  
\[
\frac{1}{1 + \exp\left(-2.97 + 0.0941 \text{ (if chronic hypertension)}\right)}
\] | 3935 |
| Van Oostwaard<sup>10</sup> 2012 | Former HDP  
And delivery 34-37 wks | • Maternal age at index  
• Chronic hypertension | Recurrent HDP any GA | Predicted risk =  
\[
\frac{1}{1 + \exp\left(-2.97 + 0.0941 \text{ (if chronic hypertension)} + 2.067\text{ (if chronic hypertension)}\right)}
\] | 5590 |

<sup>*</sup> Not available in our IPD database  
<sup>**</sup> Available IPD data that match the inclusion criteria of the development population.
All prognostic models were applied to the validation population by calculating the individual risk using the regression formulas in Table 2 for each individual in the validation population. Overall performance of the models was assessed by calibration and discrimination. Calibration was assessed graphically by plotting the predicted probabilities against the observed frequencies of the recurrence of a hypertensive disorder. The discriminative ability, being the ability of the model to distinguish women with recurrence from women without recurrence, was assessed with the c-statistic, which is for a dichotomous outcome similar to an area under the Receiver Operating Characteristic curve (AUC). A model with a c-statistic of 0.50 has no discriminative power at all (comparable to a coin flip), and a c-statistic of 1.0 reflects perfect discrimination. Although predictor-outcome associations might be generalizable to other populations, the incidence of the outcome of the model generally differs between the development and validation population, leading to either over- or underprediction of the model in the validation population. Therefore it is common practice to adjust for such differences by updating the intercepts of the original models. A correction factor was calculated by taking the logarithm of the total result of the incidence of the outcome in the validation set divided by 1 minus this incidence divided by the mean predicted risk in the validation set divided by 1 minus this mean. By adding this correction factor to the intercept of the original model, the model was updated.

Depending on the performance of the models, further adjustment of the models was considered including changing the weights of the existing predictors as a whole, also known as the slope. Adjustment of the slope (and simultaneously the intercept) was performed by adding the linear predictor (without the original intercept) of the original model as a covariate to a logistic regression model. The resulting intercept and coefficient for the linear predictor are the adjusted intercept and slope of the original model. Statistical analyses were performed using SPSS 20.0 (SPSS Inc., Chicago, IL, USA) and R version 3.0.1 (The R Foundation for Statistical Computing). The manuscript is reported in accordance to recently published guidelines.

Results

Table 3 shows the baseline characteristics and outcome of the models’ predictors in the derivation sets and in the validation populations. Differences were found between the populations for almost all predictors of the four prognostic models.
Chapter 6

### Table 3. Baseline and outcomes of the IPD and derivation sets of the prognostic models.

<table>
<thead>
<tr>
<th>Study</th>
<th>Predictors</th>
<th>Derivation set</th>
<th>IPD database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed1 2011</td>
<td>Population:</td>
<td>N = 1121</td>
<td>N = 112,975</td>
</tr>
<tr>
<td></td>
<td>- Women with former PE &lt;37 wks /HELLP (%)</td>
<td>343 (31)</td>
<td>112,975 (100)</td>
</tr>
<tr>
<td></td>
<td>- Early onset &lt; 34 wks (%)</td>
<td>151 (13)</td>
<td>42,534 (38)</td>
</tr>
<tr>
<td></td>
<td>Chronic hypertension in women with former HDP (%)</td>
<td>146 (43)</td>
<td>18,940 (17)</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>972 (87)</td>
<td>84,552 (75)</td>
</tr>
<tr>
<td></td>
<td>- White (%)</td>
<td>149 (13)</td>
<td>27,111 (24)</td>
</tr>
<tr>
<td></td>
<td>Mean blood pressures at booking</td>
<td>124 (15)</td>
<td>125 (14)</td>
</tr>
<tr>
<td></td>
<td>systolic in mmHg (SD)</td>
<td>76 (11)</td>
<td>76 (10)</td>
</tr>
<tr>
<td></td>
<td>Mean blood pressures at booking</td>
<td>478 (43)</td>
<td>32,138 (28)</td>
</tr>
<tr>
<td></td>
<td>diastolic in mmHg (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of folate (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of antihypertensive medication before subsequent pregnancy (%)</td>
<td>189 (17)</td>
<td>11,854 (10)</td>
</tr>
<tr>
<td></td>
<td>Outcome Seed1: Recurrent PE any GA (%)</td>
<td>190 (17)</td>
<td>27,246 (24)</td>
</tr>
<tr>
<td></td>
<td>Outcome Seed2: Recurrent PE &lt; 34 wks (%)</td>
<td>34 (3.0)</td>
<td>5287 (4.7)</td>
</tr>
<tr>
<td></td>
<td>- Women with former PE &lt; 34 wks (%)</td>
<td>407 (100)</td>
<td>39,348 (100)</td>
</tr>
<tr>
<td></td>
<td>Chronic hypertension (%)</td>
<td>159 (39)</td>
<td>7589 (19)</td>
</tr>
<tr>
<td></td>
<td>Mean/Median GA at former delivery in wks+days (SD/IQR)</td>
<td>29+6 (17 days)</td>
<td>31+0 (29 - 33)</td>
</tr>
<tr>
<td></td>
<td>Former SGA (%)</td>
<td>146 (36)</td>
<td>13,681 (35)</td>
</tr>
<tr>
<td></td>
<td>Mean BMI (kg/m²)</td>
<td>26 (4.7)</td>
<td>25 (22 - 29)</td>
</tr>
<tr>
<td></td>
<td>Mean Fasting blood glucose* in mmol/L (SD)</td>
<td>5.0 (0.8)</td>
<td>5.1 (0.79)</td>
</tr>
<tr>
<td></td>
<td>Outcome: Recurrent PE &lt; 34 wks (%)</td>
<td>28 (6.9)</td>
<td>3418 (8.7)</td>
</tr>
<tr>
<td>Van Oostwaard 2012</td>
<td>Population:</td>
<td>N = 312</td>
<td>N = 55,901</td>
</tr>
<tr>
<td></td>
<td>- Women with former HDP 34-37wk</td>
<td>312 (100)</td>
<td>55,901 (100)</td>
</tr>
<tr>
<td></td>
<td>Mean maternal age at index pregnancy in years (SD)</td>
<td>29 (4)</td>
<td>25 (5)</td>
</tr>
<tr>
<td></td>
<td>Chronic hypertension (%)</td>
<td>21 (7)</td>
<td>11,142 (20)</td>
</tr>
<tr>
<td></td>
<td>Outcome: Recurrent HDP any GA (%)</td>
<td>120 (38)</td>
<td>17,953 (32)</td>
</tr>
</tbody>
</table>

Data are given as percentages or as mean (SD) if normally distributed or median (IQR) if not normally distributed.
* Variable not available in the original IPD.
** Given N are after imputation. Original data are: N/10.

### Validation of existing prognostic models

**Seed1**

Recurrence of PE in the validation population for model Seed1 (former PE and delivery before 37 weeks or HELLP syndrome) occurred in 2725 (N= 11,298, 24%). Calibration of the Seed1 model showed overprediction for higher predicted risks. After the initial validation, we updated the intercept. However, since calibration was still not optimal, we also updated the slope, which improved calibration. The Seed1 model demonstrated moderate discriminative performance with a c-statistic of 0.60 (95%CI: 0.58 - 0.61). The predicted recurrence of PE at any gestational age was 57% (range: 28 - 89%) after updating intercept and slope the predicted recurrence was 24% (range: 15 - 40%).
Seed2
The prognostic model Seed2 was validated using the same validation population of 11,298 women. Recurrence and delivery before 34 weeks occurred in 529 women (4.7%). Calibration of the Seed2 model showed underprediction of lower predicted risks and slight overprediction in the higher predicted risk range. Since calibration was still not optimal after updating the intercept, we also updated the slope, after which calibration improved. The model demonstrated very good discriminative performance with a c-statistic of 0.98 (95% CI: 0.98 - 0.98). The predicted recurrence of early onset PE was 7.3% (range: 2.4 - 92%), and after updating: 4.7% (range: 0.10 - 99%).

Van Kuijk
In the validation population for model Van Kuijk (former PE < 34 weeks' gestation), recurrence of early onset PE occurred in 342 women (N= 3935, 8.7%). Calibration of the Van Kuijk model was fairly good, with small discrepancies in the extremes. Their own updated model showed worse calibration. Updating the intercept did not optimize calibration. Calibration improved after updating the slope. This new update makes the previous update of the intercept of the Van Kuijk model redundant. The model demonstrated poor discriminative performance with a c-statistic of 0.57 (95% CI: 0.53 - 0.62). The predicted recurrence of early onset PE was 13% (range: 0.95 - 50%) and after updating: 8.7% (range: 0.59 - 39%).

Van Oostwaard
Recurrence of any HDP at any gestational age in the validation population for model Van Oostwaard (previous delivery between 34 and 37 weeks of gestation) occurred in 1795 (N= 5590, 32%). Calibration of the Van Oostwaard model showed underprediction for lower predicted risks and overprediction for the higher risk range. As in the 3 other models, calibration was still not optimal after updating the intercept, which improved after also updating the slope. The prognostic model demonstrated poor discriminative performance with a c-statistic of 0.58 (95% CI: 0.56 - 0.60). The predicted recurrence of PE at any gestational age was 44% (range: 14 - 95%) and after updating intercept and slope: 32% (range: 22 - 57%).

The calibration plots of the original models and the updated Van Kuijk model are shown in Figure 2. The new calibration plots and the updated intercept and slope are shown in Figure 3. These updates did not influence the discriminative ability of
the models. Table 4 summarizes the performance and predicted recurrence of the 4 validated prognostic models.

Table 4. Performance and predicted outcome of the validated prognostic models

<table>
<thead>
<tr>
<th></th>
<th>Seed 1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>Seed 2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>Van Kuijk&lt;sup&gt;11&lt;/sup&gt;</th>
<th>Van Oostwaard&lt;sup&gt;19&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual outcome</td>
<td>24%</td>
<td>4.7%</td>
<td>8.7%</td>
<td>32%</td>
</tr>
<tr>
<td>Predicted outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- original model (range)</td>
<td>57% (28–89%)</td>
<td>7.3% (2.4–92%)</td>
<td>13% (0.95–50%)</td>
<td>44% (14–95%)</td>
</tr>
<tr>
<td>- updated model* (range)</td>
<td>24% (15–40%)</td>
<td>4.7% (0.10–99%)</td>
<td>8.7% (0.59–39%)</td>
<td>32% (22–57%)</td>
</tr>
<tr>
<td>Discriminative ability: c-statistic (95%ci)</td>
<td>0.60 (0.58–0.61)</td>
<td>0.98 (0.98–0.98)</td>
<td>0.57 (0.53–0.62)</td>
<td>0.58 (0.56–0.60)</td>
</tr>
</tbody>
</table>

* Updated intercept and slope

Discussion:

Accurate prediction of recurrence of HDP can have great advantages in counseling couples about a subsequent pregnancy. This external validation study explored the performance of 4 existing prognostic models for recurrence of HDP in a large heterogeneous cohort of women.

Main findings

In the validation of the 4 prognostic models, Seed1, Seed2, Van Kuijk and Van Oostwaard, we found c-statistics of 0.61, 0.98, 0.57 and 0.58, respectively. Mean predicted risks for recurrence were reasonably close to actual recurrence, which improved after updating intercept and slope. Calibration plots of the original models showed both under- and overprediction. After updating intercept and slope, calibration for all models was good. Discrimination was not influenced by these adjustments.
Figure 2. Calibration plots

Seed 2011

Van Kuijk 2011

Van Oostwaard 2012

Seed 2011

Van Kuijk updated

Van Kuijk updated
Figure 3. Calibration plots after updating intercept and slope

For example: the model Van Oostwaard after updating becomes as follows:

Predicted risk = \( \frac{1}{1 + \exp(-1 \times (-1.6328 + 0.3148 \times (0.0941 \times \text{maternal age} + 2.067 \times \text{chronic hypertension})))} \)
Strengths and weaknesses

Besides the challenges of including and combining data from the different studies, this IPD database proved to be valuable for the validation of 4 existing prognostic models on recurrence of HDP. This validation study fulfills a need in the growing quantity of prognostic models lacking external validation. The strength of our study is that the validation of the models was performed on a heterogeneous validation population. Weaknesses however are to be considered in the interpretation of these results. Most importantly, many studies (n= 72), identified in the previous IPD meta-analysis, could not be included in the pooled database due to various reasons and heterogeneous study methods. On a positive note, a collaboration was formed of 28 researchers of 22 studies, performed all over the world, resulting in a database with unprecedented number of women with previous HDP. Another main weakness is missing data, which originate from differences in data collection between the included studies. Missing data reduces statistical power and may introduce bias. Therefore, we used the recommended method of multiple imputation to deal with these missing values. Also, it is difficult to determine whether the performance of the models is generalizable to all included populations or whether there might be populations in which the models perform better (or worse). Unfortunately, the number of available women and cases within each individual cohort did not allow us to perform separate validation studies in each and every study population.

Not being able to obtain data for all desired studies, missing data and heterogeneity between studies seem to be a central issue for studies that aim to collect IPD from multiple studies. Nevertheless, due to the rising number of IPD studies in all fields of research, contributing to the development and validation of more prognostic models, knowledge is growing and guidance can be offered in model development, validation and reporting.

Interpretation

External validation exposed very diverse performances among the 4 existing prognostic models. The performance of the models are difficult to compare to one another, as they all focus on slightly different populations and predict different outcomes. Even though we created validation populations by selecting IPD matching the inclusion criteria of the individual prognostic models, discriminative performance was poor to moderate for the models of Seedt, Van Kuijk and Van Oostwaard. Performance of these prognostic models was lower than their previous external and/or internal validation showed. The handling of missing data, modelling continuous prognostic factors, the complexity
of the model, and model assumptions may affect performance. Poor performance may also be explained by the substantial differences between the development and validation populations regarding predictors and outcome (Table 3). Then again, the essence of external validation is to test a model in a different population. Still, even though we were unable to match the inclusion criteria for the models by Seed, one of them (Seed2) performed very well. Updating these prognostic models using this large pooled database, improves the external validity in an even wider, real-world population.

The Seed2 model performed extremely well in our IPD. The high discriminative ability likely results from the broad predicted range in this model. This way, the model discriminates well by assigning low risks to women without recurrence and high risk to women with recurrence. It was derived from a larger dataset and included more predictors than the other models. Calibration was however moderate, which improved after updating.

**Clinical applicability**

Clinical applicability of prognostic models relies on simplicity, availability of predictors, existence of preventive treatment, external validation and face validity (hesitation to rely on probabilities generated by the model). The four prognostic models validated in this IPD meta-analysis fulfill the requirements of simplicity and availability of predictors in routine care. Preventive treatments for HDP are still undetermined to a certain degree. Aspirin has been adopted as a preventive strategy, but reports present different effects of low molecular weight heparin use in pregnancy. Couples could be counseled about their individual recurrence risk for HDP, so that they can decide upon a subsequent pregnancy, based on a more rational perceived risk. Furthermore, women with a high estimated risk for recurrence can receive close surveillance. Face validity is a difficult issue to overcome, but good performance of a prognostic model in external validation should help in acknowledging its utility. External validation in this report however, does not show encouraging results for 3 out of 4 prognostic models. The model Seed2 has the most potential and deserves further evaluation in clinical practice.

A robust quality cycle is needed in the growing quantity of prognostic models, in order to eventually implement prediction of HDP in clinical use. Three out of four prognostic models showed poor prognostic performance, and although the Seed2 model definitely has potential, it is focused on a specific population, leaving room for the development of new prognostic models for recurrence of HDP. Alternatively, addition of predictors
to the models that were validated in this study may improve their performance. This IPD database can be of much value in this process, due to its size and extensive list of variables.

In conclusion, the four validated prognostic models for recurrence of disease showed varying performance. Updating with new predictors is needed to improve discriminative ability of the prognostic models. Future research is needed to evaluate clinical applicability and implementation.
References:

External validation of HDP recurrence models

Chapter 6


Contribution to Authorship:
Miriam van Oostwaard and Ewoud Schuit are the main authors, having designed the study, performed literature searches, shared data of studies that they authored and did most of the writing of the article. Ewoud Schuit performed all the statistical analyses. Josje Langenveld helped design the study, assisted in literature searches and reviewed the article. Dimitri Papatonis helped design the study, shared data and reviewed the article. Mark Brown Romano Byaruhanga, Sohinee Bhattacharya, Doris Campbell, Lucy Chappell, Francesca Chiaffarino, Isabella Crippa, Fabio Facchinetti, Sergio Ferrazzani, Enrico Ferrazzi, Ernesto Figueiró-Filho, Ingrid Gaugler-Senden, Camilla Haavaldsen, Jacob Lykke, Alfred Mbah, Vanessa Oliveira, Lucilla Poston, Christopher Redman, Read Salim, Luc Smits, Baskaran Thilaganathan, Patrizia Vergani and Jun Zhang all shared data and reviewed the article. Eric Steegers, Ben Willem Mol and Wessel Ganzevoort contributed to the design of the study, assisted and supervised with data collection and analysis and reviewed the article.