Validating a machine-learning model for first-trimester prediction of preeclampsiausing the cohort from the PREVAL study

M. M. Gil[1,2], D. Cuenca-Gómez[1,2], V. Rolle[1,3], M. Pertegal[4,5], N. Valiño[6], R. Revello[7], B. Adiego[8], M. Mendoza[9], F.S. Molina[10], B. Santacruz[1,2], Z. Ansbacher-Feldman[11], H. Meiri[12], R. Martin-Alonso[1,2], Y. Louzoun[11], C de Paco Matallana[4,5].

1. Department of Obstetrics and Gynecology, Hospital Universitario de Torrejón, Torrejón de Ardoz, Madrid, Spain.
2. Faculty of Medicine. Universidad Francisco de Vitoria, Pozuelo de Alarcón, Madrid, Spain.
3. Biostatistics and Epidemiology Platform at Instituto de Investigación Sanitaria del Principado de Asturias, Oviedo, Spain.
4. Department of Obstetrics and Gynecology, Hospital Clínico Universitario 'Virgen de la Arrixaca', El Palmar, Murcia, Spain; Institute for Biomedical Research of Murcia, IMIB-Arrixaca, El Palmar, Murcia, Spain.
5. Faculty of Medicine. Universidad de Murcia, Murcia, Spain.
6. Department of Obstetrics and Gynecology, Complejo Hospitalario Universitario A Coruña, A Coruña, Galicia, Spain.
7. Department of Obstetrics and Gynecology, Hospital Universitario Quirón, Pozuelo de Alarcón, Madrid, Spain.
8. Department of Obstetrics and Gynecology, Hospital Universitario Quirón, Pozuelo de Alarcón, Madrid, Spain.
9. Department of Obstetrics and Gynecology, Hospital Universitari Vall d'Hebrón, Barcelona, Catalonia, Spain.
10. Department of Obstetrics and Gynecology, Hospital Universitario San Cecilio, Granada, Spain and Instituto de Investigación Biosanitaria Ibs., Granada, Spain.
11. Department of Mathematics, Bar Ilan University, Ramat Gan, Israel
12. The ASPRE Consortium and TeleMarpe, Tel Aviv, Israel.

Corresponding authors:
Dr M. M. Gil
Vicerrectorado de Investigación, Facultad de Medicina, Universidad Francisco de Vitoria, Carretera Pozuelo a Majadahonda, Km 1.800, 28223 Pozuelo de Alarcón, Madrid, Spain
E-mail: mariadelmar.gil@ufv.es

Dr B. Santacruz
Dirección Médica, Hospital Universitario de Torrejón, Calle Mateo Inurria S/N, 28850 Torrejón de Ardoz, Madrid, Spain
E-mail: bsantacruz@torrejonsalud.es

Running Title: Validating first-trimester machine-learning prediction of preeclampsiain Spain

Keywords: artificial intelligence; first trimester; machine learning; mean arterial pressure; neural network; placental growth factor; posterior risk; pre-eclampsia; prior risk; uterine artery Doppler, pregnancy associated placental protein A.

CONTRIBUTION

What are the novel findings of this work?

A non-linear machine-learning model based on a combination of maternal risk factors and non-normalized first-trimester biomarkers - mean arterial blood pressure, uterine artery pulsatility index, and placental growth factor - for high accuracy prediction of early and preterm preeclampsiawas previously published. The model was developed on data from screening by the Fetal Medicine Foundation in the UK and validation was performed on a cohort from Spain without model retraining or adaptation to the population; however, adjustments were made for the analyzers used for biochemical tests because these were different in the two cohorts. The paper demonstrates how to apply a machine learning model for different populations leveraging on model training in one and its adaptation to another without specific modifications.

What are the clinical implications of this work?

The approach and methodology applied in this study will allow wider usage of advanced mathematical modelling to obtain a first-trimester prediction of preterm preeclampsiaand potentially other prenatal complications in different populations.

ABSTRACT

Objective Effective first-trimester screening for preeclampsia(PE) can be achieved with the use of a competing risks model that combines risk factors from the maternal history with multiple of the median (MoM) values of biomarkers. A new model with the use of artificial intelligence (AI) through machine-learning methods has been shown to achieve a similar performance of screening without the need for conversion of raw data of biomarkers to MoMs. We aimed to show how this model can be used across populations without specific adaptations.

Methods A machine learning model derived with the use of a fully connected neural network for first-trimester prediction of early (<34 weeks), preterm (<37 weeks), and all PE was developed and tested in a cohort of pregnant women in the UK. The model was based on maternal risk factors and mean arterial blood pressure (MAP), uterine artery pulsatility index (UtA-PI), placental growth factor (PlGF), and pregnancy-associated placental protein-A (PAPP-A). This model was applied to a dataset of 10110 singleton pregnancies examined in Spain who participated in the *First-trimester preeclampsiavalidation* (PREVAL) study, in which first-trimester screening for PE was carried out by the Fetal Medicine Foundation (FMF) competing risks model. The performance of screening was poor if no adjustments were made for the analyzer used to measure PlGF and PAPP-A, which was different in the UK and Spain. Consequently, adjustments were made for these analyzers before assessing the performance of screening by examining the area under the receiver-operating-characteristics curve (AUROC) and detection rate (DR) at 10% screen positive rate (SPR). These indices were compared to those derived from the application of the FMF competing risks model.

Results The DR at 10% SPR for early, preterm, and all PE with the machine-learning model were 84.4% (95% CI, 67.2 to 94.7%), 77.8% (95% CI, 66.4 to 86.7%), and 55.7% (95% CI, 49.0 to 62.2%), respectively, with corresponding AUROCs of 0.920 (95% CI, 0.864 to 0.920), 0.913 (95% CI, 0.882 to 0.913) and 0.846 (95% CI, 0.820 to 0.846). This performance was achieved with the use of three of the biomarkers (MAP, UtA-PI, PlGF); inclusion of PAPP-A did not provide significant improvement in DR. These results were similar to those achieved by the application of the FMF competing risks model (DR at 10% SPR 82.7%, 95% CI 69.6 to 95.8% for early PE, 72.7%, 95% CI 62.9 to 82.6% for preterm PE and 55.1%, 95% CI 48.8 to 61.4% for all PE), but did not require an adaptation of the model to the population.

Conclusion A machine learning model for first-trimester prediction of PE based on a neural network provides effective screening for PE that can be applied in different populations but before doing so it is essential to make adjustments for the analyzers used for biochemical testing.

INTRODUCTION

Machine learning methods and artificial intelligence (AI) in general are increasingly being used in many research fields including medicine. Recently, the approach was adopted for first-trimester prediction of preeclampsia (PE)[1].

In a study involving more than 60 thousand singleton pregnancies, that had been already evaluated by the Fetal Medicine Foundation (FMF) competing risks model, which combines information from maternal demographic characteristics and medical history with multiples of the median (MoMs) values of the uterine artery pulsatility index (UtA-PI), mean arterial blood pressure (MAP) and placental growth factor (PlGF), 75% of the cases of preterm PE were predicted at 10% screen positive rate (SPR)[2]. In a paper published recently, a machine learning method was applied to the same UK population to predict the risk for subsequent development of PE. The machine learning model was found to generate a similar accuracy (detection rate, DR, 75.3%, 95% confidence interval, CI, 68.9 to 81.7%, at 10% false positive rate, FPR) to that of the competing risks model without the need to adjust the values of biochemical results according to maternal and pregnancy characteristics and their conversion into MoM values[1]. Expression of values into MoMs necessitates a study of a large population and requires regular readjustment because of changing characteristics of the population and the use of different batches of the analytes[3–6]. Many centers and populations do not have large enough populations and trained laboratory personnel to continuously validate the marker data, and thus are either unable to adopt the approach or often use it automatically and lose accuracy. This is a limitation of the current FMF approach. On the other hand, machine learning and, in general, artificial intelligence methods quite often suffer from overfitting and fail to provide accurate results when a new cohort is examined[7].

We have recently validated the results from the FMF competing risks model using a cohort of 10110 singleton pregnancies as part of the *First-trimester preeclampsiavalidation* (PREVAL) study, a non-interventional prospective cohort study aiming to implement first-trimester screening of PE in Spain[8]. The DR for preterm PE achieved by this approach was 72.7% (95% CI, 62.9 to 82.6%) at 10% SPR, similar to that found in the original studies.

Our aim in this paper was to evaluate whether the machine learning model developed and tested in the UK dataset could be applied to the dataset from the PREVAL study and offer a reasonable prediction of PE. Thus, using this Spanish cohort we could reach two goals: first, validate the machine learning model developed in the UK dataset for a completely untrained dataset, and second, compare the accuracy of PE prediction for this Spanish dataset with the machine learning model using raw data to the accuracy of prediction with the FMF competing risks model where biomarker values are converted to MoMs.

METHODS

Study population

This study used the dataset collected at the PREVAL study in which women were enrolled in five different regions of Spain between September 2017 and December 2019 when they attended their routine ultrasound examination at 11^{+0}-13^{+6} weeks' gestation[8]. The study was approved by the local Research Ethics Committee at each participating hospital. Eligible women had a singleton pregnancy, with a live and non-malformed fetus at 11^{+0}-13^{+6} weeks' gestation. Women less than 18 years, those who miscarried or had pregnancy termination before 20 weeks of gestation were excluded. We also excluded those with incomplete pregnancy outcomes[8].

During the 11^{+0}-13^{+6} weeks hospital visit for prenatal care, maternal characteristics and medical history were recorded in an electronic clinical database (ViewPoint® software, GE Healthcare; Munich, Germany), and maternal weight and height were measured. Gestational age was determined from the measurement of fetal crown–rump length[9]. MAP and UtA-PI were measured according to standardized protocols[10-11]. Serum concentrations of PAPP-A and PlGF were determined using an automated device (BRAHMS KRYPTOR™ analyzer, Thermo Fisher Scientific, Hennigsdorf, Germany) during the same visit or a few days earlier, but always within the 11^{+0}-13^{+6} weeks window.

In the PREVAL study, the patients and their doctors were kept blinded to the risks of PE or the levels of biomarkers so that the use of aspirin prophylaxis was not based on such screening. However, some patients were treated with aspirin (100-150mg) according to local protocols based on maternal medical history[8].

Diagnosis of preeclampsiaand ascertainment of pregnancy outcome

Data on pregnancy outcome were obtained from hospital records, general medical practitioners, or midwives[8]. Preeclampsiawas diagnosed according to the current American College of Obstetricians and Gynecologists guidelines[12]. According to this definition, the diagnosis of PE requires the presence of new-onset hypertension (systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg) at ≥ 20 weeks' gestation or chronic hypertension and either proteinuria (≥ 300 mg/24 h or protein-to-creatinine ratio ≥ 30 mg/mmol or ≥ 2+ on dipstick testing) or evidence of renal dysfunction (serum creatinine > 97 μmol/L), hepatic dysfunction (transaminases ≥ 65 IU/L) or hematological dysfunction (platelet count < 100000/μL).

Outcome measures were early PE, preterm PE, and all PE with delivery at <34 weeks' gestation, <37 weeks, and at any gestation, respectively.

Machine learning

The input data were converted into Z-scores, based on the machine learning method developed on the dataset from the UK hospitals as was previously described, where PlGF, UtA-PI and MAP were the most important features[1]. The machine learning performed was a fully connected neural network with two hidden layers (200 and 50 neurons, respectively), and exponential linear unit as activation function. Accordingly, the normalization of the data was based on a trained machine learning dataset of 30437 patients, and was followed by an internal validation of a dataset of 10000 patients, and a tested set of 20352 patients, all from the UK[1]. The model was not further retrained or normalized on any fraction of the patients from the PREVAL study dataset of 10110 patients.

Categorical parameters were represented using one-hot encoding and were not normalized. No information on the outcome of the women in the PREVAL study dataset or on their input value distribution was available during the prediction analysis.

In the UK study measurement of PlGF and PAPP-A was with use of DELFIA Xpress system (PerkinElmer Life and Analytical Sciences, Waltham, MA, USA) or Cobas e411 system (Roche Diagnostics, Penzberg, Germany), whereas in the Spanish study, BRAHMS KRYPTOR™ analyzer was used. Analysis of data without adjustment for the analyzer provided poor results (see under Results) and therefore adjustments were necessary before the application of the UK derived model on the Spanish dataset; a simple scaling of 1.993 * PlGF – 4.225 was required to translate the BRAHMS KRYPTOR™ data to the Cobas e411 one. The scaling was performed to ensure a similar average and variance of the PlGF values over the two machines. This normalization was performed through a simple linear regression and only required a few tens of measurements.

Experimental setup

Multiple tests were performed, and in all tests, the entire 10110 patients of the PREVAL study dataset was used as both, training and test sets. In all cases, the following combinations were tested independently: (a) early PE vs. no PE, (b) preterm PE vs. no PE, and (c) any PE vs. no PE. Thus, when preterm PE was compared with no PE, term PE cases were ignored, etc.

Statistical analysis

Descriptive data were expressed as the median and interquartile range (IQR) and in proportions (absolute and relative frequencies). Comparisons between PE and non-PE groups were performed by Mann-Whitney U or $\chi 2$-test as appropriate. The level of significance was set at 0.05.

The prediction accuracy was estimated using the area under the receiver operating characteristic (AUROC) curve and the DR (recall) with 95% CI at 10% SPR. The p-values reported are the probability that the results are random. A non-parametric permutation test was used to compare the observed AUROC with the one obtained when we randomly scrambled the labels of each sample but kept its predicted score. For the general prediction, 10000 permutations of the labels were performed. All analyses were performed blindly to pregnancy outcomes. Risk calibration was assessed visually through a figure showing the observed incidence of PE against the incidence predicted from risk calculation. The plot was produced by grouping the data into bins according to the risk, and then, the observed incidence in each group was plotted against the incidence predicted by the model to the power of 10. For perfect calibration (i.e., observed incidence equals the predicted risk of PE), the slope should be 1.0, and the intercept should be 0.

The statistical software package R was used for data analyses[13]. The package Car[14] was used to clean and manage the data, the package Table1[15] was used for descriptive tables, and the package PropCIs[16] was used to calculate confidence intervals. The machine learning was performed using the sklearn[17] and pytorch[18] in Python[19].

RESULTS

Study population

The study population comprised 10110 singleton pregnancies, including 32 (0.32%) cases of early PE, 72 (0.7%) cases of preterm PE, and 230 (2.27%) cases of all PE. The characteristics of the population are summarized in Table 1.

Performance of the machine learning model for the prediction of pre-eclampsia

The first results, derived from direct application of the UK-derived model without any adjustments for the analyzed for PlGF were poor[1]; for example, the DR, at 10% SPR, for preterm PE of only 41.7% (95% CI 30.2 to 53.9%). The performance improved after appropriate adjustments for the analyzer used.

Table 2 provides the results of the application of the UK-based model adjusted for analyzer for PlGF for different combinations of maternal risk factors with biomarkers. The best performance of screening was achieved by a combination of maternal factors with MAP, UtA-PI, and PlGF. The DR at 10% SPR for early, preterm, and all PE with the machine-learning model were 84.4% (95% CI 67.2 to 94.7%), 77.8% (95% CI 66.4 to 86.7%), and 55.7% (95% CI 49.0 to 62.2%), respectively, with corresponding AUROCs of 0.920 (95% CI 0.864 to 0.920), 0.913 (95% CI 0.882 to 0.913) and 0.846 (95% CI 0.820 to 0.846). These results were similar to those achieved by the application of the FMF competing risks model (DR at 10% SPR 82.7%, 95% CI 69.6 to 95.8% for early PE, 72.7%, 95% CI 62.9 to 82.6% for preterm PE and 55.1%, 95% CI 48.8 to 61.4% for all PE).

The addition of PAPP-A to the model based on maternal characteristics, MAP, and UtA-PI improved the DR, this improvement was lower than that provided by adding PlGF instead and including PAPP-A when PlGF was already in the algorithm, which did not improve the screening performance.

The performance of screening of the UK-based model, adjusted for the analyzer for PlGF, in the Spanish dataset was similar to that obtained in the UK dataset (Supplementary Table 1).

Calibration plots (Figure 2) showed a moderate correlation between predicted and observed cases of preterm PE, with a slope of 0.702 (95% CI, 0.445 to 0.960) and an intercept of -0.723 (95% CI, -1.636 to 0.188).

DISCUSSION

Main findings

In this study we have shown that, after applying appropriate adjustments for the analyzer used in the new data, the DR for early (84.4%, 67.2 to 94.7%), preterm (77.8%, 66.4 to 86.7%) and all PE (55.7%, 49.0 to 62.2%) at 10% SPR and AUROC achieved by the machine learning model is similar to first, that achieved by the FMF competing risk model in the same population and, second, that achieved in the original dataset where the model was developed. These results demonstrate that the model can be applied from one dataset to another and yield very similar predictions as long as adjustments are made for biomarkers using different platforms. This model, which uses raw data for the different biomarkers would avoid the need for MoM conversion.

Additionally, we have provided further evidence that, once PlGF is incorporated in the algorithm, the inclusion of PAPP-A does not provide any significant improvement in the screening performance.

Comparison with previous studies and clinical implications

First-trimester prediction of preterm PE is important because the ASPRE trial demonstrated that daily treatment of the high-risk group with low-dose aspirin (150 mg/day from 12 to 36 weeks' gestation) reduces the rate of early PE with delivery < 32 weeks by about 90% and preterm PE by about 60%[20-22] Consequently, implementing an accurate method of screening for such condition has been recommended by many societies[23,24].

The predictive performance for preterm PE using this machine learning model was similar to that achieved by the competing risks model with no statistical significance diferences both in terms of overall performance and calibration plots[2,8,20,25,26]. Since the machine learning approach uses raw biomarker data without any conversion into MoMs, implementation of screening would be simplified in many settings, particularly those where the number of screened women is too small to allow for accurate assessment and adjustment of MoM values. This model has already been made available online (https://pepred.math.biu.ac.il/Home.) and it is ready for clinical use.

Moreover, the increasing evidence that PAPP-A does not substantially contribute to the performance of PE screening supports the recommendation that this screening should never be offered under the umbrella of first-trimester screening of aneuploidies but on its own. Even in those settings where PlGF analysis is not possible and a different combination of biomarkers is chosen, PE screening should be discussed independently of aneuploidy screening.

Finally, the assumption is that after each validation, the validated model can be added to the existing dataset and the model can be retrained, helping to build better prediction accuracy each time. This methodology will allow wider usage of advanced mathematical modelling not only for PE screening but, potentially, for other prenatal complications in different populations.

Limitations and strengths

The main limitation of the study is the small number of cases of early and preterm PE with consequent difficulty to achieve statistical significance in differences between models. However, given that the differences consisted only of a few point percentages, it is unlikely that a larger number would have yielded different results. Another limitation was that about 6% of our patients were taking aspirin. Since we know that aspirin prophylaxis reduces the incidence of preterm PE by about 60%[21,22]. Since we have not adjusted by the use of aspirin, it is possible that the incidence of PE is slightly higher than reported in the PREVAL study and that some cases that, without aspirin were deemed to develop PE, have been converted into

false positives, worsening the screening performance. However, this is unlikely to have introduced a large bias because the number of cases taking aspirin was small.

The main strength of the study relates to the large number of non-affected pregnancies, which has allowed us to accurately adapt the PlGF values from one analyser to another, enabling the algorithm to be used by the three platforms validated for first-trimester screening of PE. To maintain this high-level performance, a new platform would require retraining of the model. However, this requires a very small number of measurements (<100) to get a rough estimate of the average of the PlGF results. Additionally, neural networks achieve higher discrimination than other traditional methods but are complex models and therefore, more difficult to interpret, unlike regression models which provide better understandable coefficients. However, while the regression model requires normalization (MoM conversion), the ML based method does not require it, making them simpler for clinical practice. Finally, we have also been able to identify differences between populations, demonstrating that the model performs equally well.

Conclusion

A machine learning model for first-trimester prediction of PE based on a neural network provides effective screening for PE that can be applied in different populations but before doing so it is essential to make adjustments for the analyzers used for biochemical testing.

ACKNOWLEDGMENTS

This study is part of the Ph.D. thesis of Diana Cuenca. Gil MM was awarded a Leonardo grant from the BBVA Foundation to conduct this study. This study was also supported by a grant from Instituto de Salud Carlos III (ISCIII) PI18/01749. iMaterna Foundation (Registry No: 2148) provided smaller grants to support the study, Thermo Fisher Scientific (Hennigsdorf, Germany) provided the reagents and instruments to perform the analysis of serum biomarkers and General Electrics (GE Medical Systems, Zipf, Austria) via Health Net Connections (HNC, Spain) provided the software (ViewPoint®) to collect the data. We are also greatfull to the Synlab International GmbH, for providing the human and technical resources to analyse some of the blood samples of the study. None of these bodies had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. We are particularly grateful for the generous contribution of the patients, the laboratory personnel involved in the analysis of samples and the nurses / auxiliary nurses involved in pregnancy care and the collaboration of Biobank Network of the Region of Murcia, BIOBANC-MUR (reg.number B.0000859), supported by the "Instituto de Salud Carlos III (proyecto PT20/00109), by "Instituto Murciano de Investigación Biosanitaria Virgen de la Arrixaca, IMIB" and by "Consejeria de Salud de la Comunidad Autónoma de la Región de Murcia" and the Biobanco A Coruña (National Registry of Biobanks B.0000796) integrated in the Platform ISCIII Biobanks and Biomodels (PT20/00128) for their collaboration. Z.A. was sponsored by Bar Ilan University Graduate program.

Finally, the authors are grateful to Professor Kypros H Nicolaides for helping with the study design, discussion and for making the UK dataset available for the comparative analysis.

REFERENCES

1. Ansbacher-Feldman Z, Syngelaki A, Meiri H, Cirkin R, Nicolaides KH, Louzoun Y. Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers. *Ultrasound Obstet Gynecol* 2022; **60**: 739-745.
2. Tan MY, Syngelaki A, Poon LC, Rolnik DL, O'Gorman N, Delgado JL, Akolekar R, Konstantinidou L, Tsavdaridou M, Galeva S, Ajdacka U, Molina FS, Persico N, Jani JC, Plasencia W, Greco E, Papaioannou G, Wright A, Wright D, Nicolaides KH. Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. *Ultrasound Obstet Gynecol* 2018; **52**: 186-195.
3. Tayyar A, Guerra L, Wright A, Wright D, Nicolaides KH. Uterine artery pulsatility index in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 689–697.
4. Wright A, Wright D, Ispas CA, Poon LC, Nicolaides KH. Mean arterial pressure in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 698–706.
5. Wright D, Silva M, Papadopoulos S, Wright A, Nicolaides KH. Serum pregnancy-associated plasma protein-A in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **46**: 42–50.
6. Tsiakkas A, Duvdevani N, Wright A, Wright D, Nicolaides KH. Serum placental growth factor in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 591–598.
7. https://www.v7labs.com/blog/overfitting. Accessed on April 28, 2023
8. Cuenca-Gómez, D., De Paco Matallana, C., Rolle, V., Valiño, N., Revello, R., Adiego, B., Mendoza, M., Molina, F.S., Carrillo, M.P., Delgado, J.L., Wright, A., Santacruz, B. and Gil, M.M. (2023), Performance of first-trimester combined screening of preterm pre-eclampsia: results from cohort of 10 110 pregnancies in Spain. Ultrasound Obstet Gynecol. Accepted Author Manuscript. https://doi.org/10.1002/uog.26233
9. Robinson HP, Fleming JE. A critical evaluation of sonar ‘‘crown–rump length’’ measurements. *Br J Obstet Gynaecol* 1975; **82**: 702–710.
10. Poon LC, Zymeri NA, Zamprakou A, Syngelaki A, Nicolaides KH. Protocol for measurement of mean arterial pressure at 11- 13 weeks' gestation. *Fetal Diagn Ther* 2012; **31**: 42–48.
11. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol* 2007; **30**: 742-9.
12. American College of Obstetricians and Gynecologists, and the Task Force on Hypertension in Pregnancy. Hypertension in Pregnancy. *Obstet Gynecol* 2013; **122**: 1122–1131.
13. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2011.
14. Fox J, Sandford W. An R Companion to Applied Regression. Third Edit. Thousand Oaks; 2019. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
15. Benjamin R. (2020). table1: Tables of Descriptive Statistics in HTML. R package version 1.2. https://CRAN.R-project.org/package=table1
16. Ralph S. (2018). PropCIs: Various Confidence Interval Methods for Proportions. R package version 0.3-0. https://CRAN.R-project.org/package=PropCIs
17. Pedregosa F, Varoquaux, Ga"el, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research 2011; 12: 2825–2830.
18. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, Inc.; 2019. p. 8024–35. Available from: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

19. Van Rossum G, Drake Jr FL. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam; 1995.
20. O'Gorman N, Wright D, Syngelaki A, Akolekar R, Wright A, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation. *Am J Obstet Gynecol* 2016; **214**: 103.e1–12.
21. Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, de Paco Matallana C, Akolekar R, Cicero S, Janga D, Singh M, Molina FS, Persico N, Jani JC, Plasencia W, Papaioannou G, Tenenbaum-Gavish K, Meiri H, Gizurarson S, Maclagan K, Nicolaides KH. Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. *N Engl J Med* 2017; **377**: 613–622.
22. Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis*. Am J Obstet Gynecol* 2018; **218**: 287–293.e1
23. Magee LA, Brown MA, Hall DR, Gupte S, Hennessy A, Karumanchi SA, Kenny LC, McCarthy F, Myers J, Poon LC, Rana S, Saito S, Staff AC, Tsigas E, von Dadelszen P. The 2021 International Society for the Study of Hypertension in Pregnancy classification, diagnosis & management recommendations for international practice. *Pregnancy Hypertens* 2021; ***27***: 148–169.
24. Poon LC, Shennan A, Hyett JA, Kapur A, Hadar E, Divakar H, McAuliffe F, da Silva Costa F, von Dadelszen P, McIntyre HD, Kihara AB, Di Renzo GC, Romero R, D'Alton M, Berghella V, Nicolaides KH, Hod M. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention. *Int J Gynaecol Obstet* 2019; **145**: 1–33.
25. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. *Am J Obstet Gynecol* 2020; **223**: 12–23.e7.
26. Wright D, Tan MY, O'Gorman N, Poon LC, Syngelaki A, Wright A, Nicolaides KH. Predictive performance of the competing risk model in screening for preeclampsia. *Am J Obstet Gynecol* 2019; **220**: 199.e1–13.

Table 1. Characteristics of the study population.

Characteristic	Non-PE (*n* = 9880)	PE (*n* = 230)	p-value
Maternal age (years)	33.9 [30.1, 36.9]	34.5 [31.2, 38.4]	0.0324
Body mass index (kg/m2)	24.0 [21.7, 27.2]	27.1 [23.7, 31.7]*	<0.001
GA at screening (days)	89.0 [86.0, 91.0]	89.0 [86.0, 92.0]	0.1442
Race			
White	9604 (97.2%)	213 (92.6%)*	<0.001
Black	70 (0.709%)	8 (3.48%)*	<0.001
South Asian	7 (0.0709%)	0 (0%)	1
East Asian	21 (0.213%)	1 (0.435%)	0.788
Mixed	178 (1.80%)	8 (3.48%)	0.032
Medical history			
Chronic hypertension	82 (0.830%)	24 (10.4%)*	<0.001
DM Type I	45 (0.455%)	6 (2.61%)*	0.426
DM Type II	20 (0.202%)	3 (1.30%)*	0.055
SLE/APS	52 (0.526%)	1 (0.435%)	0.715
Smoker	1137 (11.5%)	23 (10.0%)	0.682
Family history of PE	273 (2.76%)	19 (8.26%)*	<0.001
Method of conception			
Spontaneous	9091 (92.0%)	194 (84.3%)*	0.012
Assisted	789 (7.99%)	36 (15.7%)*	0.012
Obstetric history			
Nulliparous	4955 (50.2%)	142 (61.7%)*	0.004
Parous, no previous PE	4782 (48.4%)	61 (26.5%)*	<0.001
Parous, previous PE	143 (1.45%)	27 (11.7%)*	<0.001
Aspirin use	534 (5.40%)	30 (13.0%)*	<0.001
Biomarker			
MAP (mmHg)	0.990 [0.936, 1.05]	1.06 [1.00, 1.12]*	<0.001
UtA-PI	1.01 [0.818, 1.22]	1.19 [0.987, 1.41]*	<0.001
PlGF (pg/mL)	0.785 [0.579, 1.04]	0.568 [0.391, 0.818]*	<0.001
PAPP-A (IU/L)	0.917 [0.579, 1.40]	0.817 [0.426, 1.30]*	0.0515

Data are given as *n* (%) or median (interquartile range). APS, antiphospholipid syndrome; DM, diabetes mellitus; GA, gestational age; MAP, mean arterial pressure; PAPP-A, pregnancy-associated placental protein-A; PE, pre-eclampsia; PlGF, placental growth factor; SLE, systemic erythematosus lupus; UtA-PI, uterine artery pulsatility index.

Table 2. Performance of screening for preeclampsia at <34 weeks, <37 weeks and at any gestational age achieved by the machine learning model in comparison to that of the competing risks model applied in the study population of the PREVAL study.

Screening model	Current study		PREVAL study	
	DR (95% CI) at 10% SPR	AUROC (95% CI)	DR (95% CI) at 10% SPR	AUROC (95% CI)
Preeclampsia <34 weeks				
Maternal factors	31.2 (16.1 to 50.0)	0.718 (0.623 to 0.718)	35.7 (17.5 to 53.9)	0.69 (0.58 to 0.79)
+ MAP + UtA-PI	68.8 (50.0 to 83.9)	0.881 (0.813 to 0.881)	64.4 (47.3 to 81.6)	0.87 (0.81 to 0.94)
+ MAP + UtA-PI + PAPP-A	75.0 (56.6 to 88.5)	0.889 (0.823 to 0.889)	68.2 (52.1 to 84.4)	0.86 (0.79 to 0.93)
+ MAP + UtA-PI + PlGF	84.4 (67.2 to 94.7)	0.920 (0.864 to 0.920)	82.7 (69.6 to 95.8)	0.94 (0.90 to 0.98)
+ MAP + UtA-PI + PlGF + PAPP-A	81.2 (63.6 to 92.8)	0.921 (0.865 to 0.921)	82.5 (69.8 to 95.2)	0.94 (0.90 to 0.97)
Preeclampsia <37 weeks				
Maternal factors	36.0 (22.0 to 42.5)	0.714 (0.655 to 0.714)	37.5 (26.8 to 48.2)	0.71 (0.65 to 0.77)
+ MAP + UtA-PI	66.7 (54.6 to 77.3)	0.880 (0.842 to 0.880)	64.2 (53.5 to 74.9)	0.87 (0.83 to 0.91)
+ MAP + UtA-PI + PAPP-A	69.4 (57.5 to 79.8)	0.870 (0.826 to 0.870)	66.5 (55.8 to 77.2)	0.87 (0.83 to 0.91)
+ MAP + UtA-PI + PlGF	77.8 (66.4 to 86.7)	0.913 (0.882 to 0.913)	72.7 (62.9 to 82.6)	0.92 (0.89 to 0.95)
+ MAP + UtA-PI + PlGF + PAPP-A	77.8 (66.4 to 86.7)	0.912 (0.880 to 0.912)	71.6 (61.7 to 81.5)	0.92 (0.88 to 0.95)
All preeclampsia				
Maternal factors	34.8 (28.6 to 41.3)	0.734 (0.701 to 0.734)	36.6 (30.2 to 43.0)	0.73 (0.69 to 0.76)
+ MAP + UtA-PI	51.7 (45.1 to 58.4)	0.829 (0.802 to 0.829)	51.0 (44.7 to 57.3)	0.82 (0.79 to 0.85)
+ MAP + UtA-PI + PAPP-A	51.3 (44.6 to 57.9)	0.829 (0.801 to 0.829)	51.4 (45.0 to 57.8)	0.82 (0.79 to 0.85)
+ MAP + UtA-PI + PlGF	55.7 (49.0 to 62.2)	0.846 (0.820 to 0.846)	55.1 (48.8 to 61.4)	0.83 (0.81 to 0.86)
+ MAP + UtA-PI + PlGF + PAPP-A	54.8 (48.1 to 61.3)	0.848 (0.822 to 0.848)	55.8 (49.5 to 62.1)	0.83 (0.81 to 0.86)

DR, detection rate; SPR, screening positive rate; AUROC, area under the receiving operator characteristic; MAP, mean arterial pressure; UtA-PI, uterine artery pulsatility index PAPP-A, pregnancy-associated plasma protein-A; PlGF, placental growth factor.

Table 1S. Performance of screening for preeclampsiaat <34 weeks, <37 weeks and at any gestational age achieved by the machine learning model in comparison to that of the original model in the UK data.

Screening model	**Current study**		**ANSBACHER-FELDMAN study**	
	DR (95% CI) at 10% SPR	**AUROC (95% CI)**	**DR (95% CI) at 10% SPR**	**AUROC (95% CI)**
Preeclampsia<34 weeks				
Maternal factors	31.2 (16.1 to 50.0)	0.718 (0.623 to 0.718)	54.6 (42.0 to 67.4)	0.75 (0.63 to 0.83)
+ MAP + UtA-PI	68.8 (50.0 to 83.9)	0.881 (0.813 to 0.881)	73.7(57.9 to 87.6)	0.93 (0.88 to 0.97)
+ MAP + UtA-PI + PAPP-A	75.0 (56.6 to 88.5)	0.889 (0.823 to 0.889)	72.7(57.9 to 87.6)	0.94 (0.89 to 0.97)
+ MAP + UtA-PI + PlGF	84.4 (67.2 to 94.7)	0.920 (0.864 to 0.920)	81.8 (66.1 to 97.6)	0.95 (0.91 to 0.97)
+ MAP + UtA-PI + PlGF + PAPP-A	81.2 (63.6 to 92.8)	0.921 (0.865 to 0.921)	81.8(66.1 to 97.6)	0.95 (0.90 to 0.97)
Preeclampsia<37 weeks				
Maternal factors	36.0 (22.0 to 42.5)	0.714 (0.655 to 0.714)	43.0 (36.0 to 50.1)	0.75 (0.67 to 0.81)
+ MAP + UtA-PI	66.7 (54.6 to 77.3)	0.880 (0.842 to 0.880)	65.1(56.4 to 73.8)	0.90 (0.86 to 0.94)
+ MAP + UtA-PI + PAPP-A	69.4 (57.5 to 79.8)	0.870 (0.826 to 0.870)	65.1(56.4 to 73.8)	0.91 (0.86 to 0.94)
+ MAP + UtA-PI + PlGF	77.8 (66.4 to 86.7)	0.913 (0.882 to 0.913)	73.3 (64.0 to 82.5)	0.91 (0.86 to 0.94)
+ MAP + UtA-PI + PlGF + PAPP-A	77.8 (66.4 to 86.7)	0.912 (0.880 to 0.912)	73.3 (64.0 to 82.5)	0.91 (0.86 to 0.95)
All pre-eclampsia				
Maternal factors	34.8 (28.6 to 41.3)	0.734 (0.701 to 0.734)	31.5 (28.1 to 34.8)	0.72 (0.69 to 0.75)
+ MAP + UtA-PI	51.7 (45.1 to 58.4)	0.829 (0.802 to 0.829)	43.8 (39.9 to 47.7)	0.79 (0.77 to 0.83)
+ MAP + UtA-PI + PAPP-A	51.3 (44.6 to 57.9)	0.829 (0.801 to 0.829)	44.5 (40.6 to 48.5)	0.80 (0.76 to 0.83)
+ MAP + UtA-PI + PlGF	55.7 (49.0 to 62.2)	0.846 (0.820 to 0.846)	47.0 (42.9 to 51.1)	0.80 (0.77 to 0.84)
+ MAP + UtA-PI + PlGF + PAPP-A	54.8 (48.1 to 61.3)	0.848 (0.822 to 0.848)	47.3 (43.3 to 51.4)	0.80 (0.77 to 0.84)

DR, detection rate; SPR, screening positive rate; AUROC, the area under the receiving operator characteristic; MAP, mean arterial pressure; UtA-PI, uterine artery pulsatility index PAPP-A, pregnancy-associated plasma protein-A; PlGF, placental growth factor.

Figure 1. Receiver operating characteristics curves of the FMF competing risks model (interrupted lines) and the machine learning model (solid lines) to predict preeclampsiaat any gestational age (in red) and preeclampsia< 37 weeks (in black).

FIGU **Figure 2.** Calibration of the model. The y-axis represents the observed incidence of preterm pre-eclampsia, and the x-axis represents the predicted incidence of preterm preeclampsia (screen-positive cases). Grey numbers on top of each group are the observed number of preeclampsia cases delivered preterm and black numbers are the observed number of cases delivered without preeclampsiafor each bin of predicted risk.