medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Deep learning-based prognosis models accurately predict the time to delivery among preeclampsia patients using health records at the time of diagnosis

Xiaotong Yang, MS[1]; Hailey K Ballard, BS[2]; Aditya D Mahadevan, BS[4,5]; Ke Xu, MS[2,5]; David G Garmire, PhD[6]; Elizabeth S Langen, MD[7]; Dominick J Lemas, PhD[2,3,5]; Lana X Garmire, PhD[1*]

1 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America

2 Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, United States of America

3 Department of Obstetrics & Gynecology, University of Florida College of Medicine, Gainesville, Florida

4 Department of Physiology and Aging, University of Florida, Gainesville, FL, United States

5 Center for Research in Perinatal Outcomes, University of Florida, Gainesville, FL, United States

6 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, United States of America

7 Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, Michigan, United States of America

*: Corresponding author:

Lana Garmire

lgarmire@umich.edu

Room 3366, Building 520, NCRC,

1600 Huron Parkway, Ann Arbor, MI 48105

Office phone: (734) 615-0514

word count: 5067

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abbreviations:

PE: preeclampsia

sPE: Preeclampsia with severe features

EOPE: early-onset preeclampsia

LOPE: late-onset preeclampsia

EHR: electronic health record

SBP: systolic blood pressure

DBP: diastolic blood pressure

RR: respiratory rate

HELLP: hemolysis, elevated liver enzymes, low platelet count

AST: aspartate transaminase

PI: prognosis score

UM: University of Michigan

UF: University of Florida

ICD-10: The International Classification of Diseases, Tenth Revision

MAP: mean arterial pressure

UtA-PI: uterine artery pulsatility index

PLGF: placental growth factor

ACOG: American College of Obstetricians and Gynecologist

AUROC: Area Under the Receiver-Operator curve

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background

Preeclampsia (PE) is one of the leading factors in maternal and perinatal mortality and morbidity worldwide with no known cure. Delivery timing is key to balancing maternal and fetal risk in pregnancies complicated by PE. Delivery timing of PE patients is traditionally determined by closely monitoring over a prolonged time. We developed and externally validated a deep learning models that can predict the time to delivery of PE patients, based on electronic health records (EHR) data by the time of the initial diagnosis, in the hope of reducing the need for close monitoring.

Method

Using the deep-learning survival model (Cox-nnet), we constructed time-to-delivery prediction models for all PE patients and early-onset preeclampsia (EOPE) patients. The discovery cohort consisted of 1,533 PE cases, including 374 EOPE, that were delivered at the University of Michigan Health System (UM) between 2015 and 2021. The validation cohort contained 2,172 PE cases (547 EOPE) from the University of Florida Health System (UF) in the same time period. We built clinically informative baseline models from 45 pre-diagnosis clinical variables that include demographics, medical history, comorbidity, PE severity, and initial diagnosis gestational age features. We also built full models from 60 clinical variables that include additional 15 lab tests and vital signs features around the time of diagnosis.

Results

The 7-feature baseline models on all PE patients reached C-indices of 0.74 and 0.73 on UM hold-out testing and UF validation dataset respectively, whereas the 12-feature full model had improved C-indices of 0.79 and 0.74 on the same datasets. For the more urgent EOPE cases, the 6-feature baseline model achieved C-indices of 0.68 and 0.63, and its 13-feature full model counterpart reached C-indices of 0.76 and 0.67 in the same datasets.

Conclusions

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We successfully developed and externally validated an accurate deep-learning model for time-to-delivery prediction among PE patients at the time of diagnosis, which helps to prepare clinicians and patients for expected deliveries.

Introduction

Preeclampsia (PE) is a pregnancy complication affecting 2% to 8% of all pregnancies worldwide and is a leading cause of maternal, fetal, and neonatal mortality and morbidity[1,2]. PE is defined by new-onset hypertension after 20 weeks of gestation and the presence of proteinuria, and/or other signs of end-organ damage. PE is a diverse syndrome with various subtypes along the spectrum of gestational hypertensive disorders[3]. It can be divided into early-onset PE (diagnosed before 34 weeks of pregnancy) or late-onset PE (diagnosed after 34+0 weeks of pregnancy); PE with severe features (sPE) or PE without severe features[4,5]. Failure to properly manage PE can lead to a wide variety of severe maternal and neonatal adverse outcomes according to the iHOPE study, while the only known cure for PE is delivery of the placenta[6–8]. Although earlier delivery can significantly reduce the risk of maternal adverse outcomes, it is associated with increased neonatal unit admission among preterm patients. This, especially in cases of EOPE [9], creates a dilemma as earlier delivery can potentially prevent severe morbidities including maternal seizure, stroke, organ dysfunction, and intrauterine fetal demise, but may lead to premature birth and subsequent neonatal complications[10,11]. To balance the risks to both mother and baby, current clinical management of PE includes supportive blood pressure management and prophylaxis for maternal seizures, and a two-dose intramuscular course of betamethasone to augment fetal lung maturation[12].

Generally, delivery is recommended for PE patients with more than 37+0 weeks of gestation and for severe PE patients with more than 34+0 weeks of gestation[12]. In reality, the delivery timing is a more complex problem, clinicians need to consider both the fetal development, maternal and fetal risk of

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

complications, and availability of ICU resources when deciding on delivery timing, particularly among challenging EOPE cases[12,13]. The decision of delivery is usually made after close monitoring and extensive testing on preeclampsia patients over a prolonged time, which may not be easily accessible and affordable to all patients (particularly those in rural areas or under-developed countries). In addition, current risk assessment tools focus on maternal risk prediction but not the overall delivery urgency considering both moms and fetuses. FullPIERS, miniPIERS and PREP-S are well-established and externally validated models to predict the maternal risk of adverse outcomes among PE patients, in the hope of assisting delivery decisions[14–19]. These tools are recommended by some, but not all national guidelines[12,13]. Most of these tools only predict maternal risks, however, clinicians need to consider both maternal and neonatal outcomes when deciding when to deliver. A patient at 34 weeks of gestation would have very different delivery timing compared to a patient at 37 weeks of gestation, even if they have the same risk of adverse outcomes. It is of great importance to directly and precisely predict the time to delivery as early as the first diagnosis of PE, which allows the clinicians to assess the delivery urgency early on and to help them better prioritize resources and treatment, particularly for those doctors practicing in rural or under-developed countries. Additionally, the aforementioned risk predictor models do not assess the risk from baseline features, such as the patient's race, social status, lifestyle, and other comorbidities, which may also have influences on delivery timing.

Toward this goal, we developed and externally validated the first deep learning model to predict patient delivery time after the initial diagnosis of PE using electronic health records (EHR) data. We utilized the state-of-the-art deep learning-based prognosis prediction model, Cox-nnet (version 2), which we previously developed [20–22]. Cox-nnet methods had previously consistently shown excellent predictive performances under a variety of conditions, including on EHR data[20]. Our objectives were: (1) to predict the time to delivery at the first diagnosis of PE for all PE patients and an EOPE sub-cohort, by constructing and validating deep-learning models utilizing EHR data; and (2) to assess the quantitative

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contributions of critical EHR features informative of delivery time among PE patients, including those EOPE patients.

Methods

Data Source

We obtained the discovery cohort of the Precision Health Initiative of Michigan Medicine (UM), the academic healthcare system of the University of Michigan[23]. Data usage was approved by the Institutional Review Board (IRB) of the University of Michigan Medical School (HUM#00168171). We obtained the validation cohort from the Integrated Data Repository database at the University of Florida (UF). Data usage was approved by IRB of the University of Florida (#IRB201601899). In both cohorts, we extracted all obstetric records with at least one PE diagnosis between 2015 to 2021 based on ICD-10 diagnosis codes (**Supplementary Table 1**). We excluded patients with the following conditions: Hemolysis, Elevated Liver Enzymes, and Low Platelet (HELLP) syndrome and eclampsia, for which iatrogenic delivery is ubiquitously induced within 48 hours of diagnosis despite fetal condition; chronic hypertension with superimposed PE, whose onset may occur before week 20 and with no clear definitions in the United States[24]; and postpartum PE, which is only developed after delivery. We also removed patients transferred from other institutions by deleting patients with no visit record within 180 days before the first diagnosis of PE to ensure the accuracy of the initial diagnosis time of PE. The resulting discovery cohort consisted of 1,533 PE cases (including 374 EOPE cases) and the validation cohort contained 2,172 PE cases (including 547 EOPE).

Fully connected Cox-nnet neural network models

We constructed all models using the Cox-nnet v2 algorithm (**Supplementary Figure 1**)[20]. In this study, we adopted the model to predict the time between PE diagnosis to delivery. To ensure the stability of the

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

models, we divided the discovery dataset into a training set (80%) and a hold-out testing set (20%) and applied 5-fold cross-validation on the training set.

EHR Feature Engineering

We extracted all available features from UM Precision Health Initiative EMR data. We developed 4 models to predict the time to delivery of PE patients: PE baseline, PE full, EOPE baseline and EOPE full models. As suggested by clinicians, the initial baseline models include demographics, medical history, comorbidities, the severity of PE, pregnancy and fetal development characteristics. The full model incorporated all features from the baseline model, with additional lab results and vital signs commonly collected within 5 days before the initial diagnosis of PE (**Supplementary Figure 2A**). EOPE models were built and tested using the same features on patients with PE onset time before 34 weeks of gestation. Features with low powers and high correlation were removed to ensure model accuracy.

Pregnancy characteristics included parity, number of fetuses, gestational age, PE severity at initial diagnosis, and history of preterm birth, c-section, abruption, etc. Fetal development includes poor fetal growth according to the associated ICD code(O36.59). Other comorbidities were grouped into 29 categories using the Elixhauser Comorbidity Index[1]. The observational window for lab results and vital signs was 5 days before the day of the initial PE diagnosis. Only the first results of repeated lab tests were used to avoid intervention/treatment effects. Summary statistics of systolic blood pressure (SBP), diastolic blood pressure (DBP), and respiratory rate (RR) measures were included (max, min, mean, standard deviation), as done in previous work[2]. We removed features with high missing proportions (over 20%) and sparse features with fewer than 10 non-zero values. Highly correlated variables were identified using the variance inflation factor (VIF) and removed one at a time until all features had a VIF below 3 to avoid multicollinearity. The remaining missing values were imputed using the PMM algorithm from R package “mice”. All numerical features were scaled by dividing their root mean square. Numeric features

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with skewness above 3 were log-transformed. As a result, 60 features were kept for initial analysis (**Supplementary Table 2, Supplementary Figure 2B).**

Reduced feature representation from the Cox-nnet models

To derive a subset of clinically significant and easily interpretable features, we reduced Cox-nnet features based on both their importance scores and significance levels. To do so, we first selected the top 15 (25% of total features) most important features based on their average permutation importance scores generated by Cox-nnet models. Permutation important scores provide more stable results than other feature selection methods on this dataset, including stepwise selection, lasso regularization, and random forest feature selection[25]. Then we calculated the log-rank p-value for the 15 features individually and selected the significant ones. We also conducted the ANOVA test on the remaining features to ensure their powers (**Supplementary Table 3**). We rebuilt the clinically informative Cox-nnet models with the reduced set of features, the same way as the models using all initial input features.

Model evaluation

We evaluated the cross-validation, hold-out test, and validation results of each model using Harrel's concordance index (C-index). The C-index evaluates the accuracy of predicted events by comparing their relative order to the order of actual events. It is frequently used to assess survival predictions[26]. The reported C-indices in the training data are the repeated results of the 5-fold cross-validation C-indices on the training sets. To enhance the interpretation of the prognosis prediction, we also stratified patients into high, medium, and low-urgency groups based on the predicted results plotted the Kalper-Meier (KM) curves of time-to-delivery in each group and reported the log-rank p-values. The log-rank test, on the other hand, compares the survival distribution between patient groups, assuming no differences in survival exist[27(p4)]. Additionally, we used each clinically informative and reduced model result to predict the chances of patients delivering within 2 days, 7 days, and 14 days and obtained the AUROC (area under the receiver operating curve) for each task.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Interactive Web Application for Easy Model Validation

To disseminate the models for public use, we containerized the pre-trained Cox-nnet model into a Docker-based web application using R shiny[28]. This allows the users to access the models easily through a local web interface and get prediction results quickly. This app contains two main panels: the individual prediction panel and the group prediction panel. Using pre-trained models, the individual prediction panel calculates the prognosis index (PI) score of a single new patient, marking its positions and percentiles in a distribution plot of PIs within the UM discovery cohort. The group panel takes in a group of new patients and returns predicted PIs and percentiles of their PIs in a table. The shiny app is available at http://garmiregroup.org/PE-prognosis-predictor/app

External Validation using UF data

We validated the reduced models on a large external EHR dataset from the University of Florida. We extracted and processed the same features included in the baseline, full, EOPE-baseline and EOPE-full model (see Methods). The authors uploaded cleaned UF data to the shiny app described above, and the app automatically produced predicted values using the packaged models trained on UM dataset. The development and validation strictly followed the TRIPOD checklist(**Supplementary Table 4**).

Estimate time-to-deliver using maternal risk of adverse outcomes calculated from the fullPIERS model

The fullPIERS model is a model to predict the maternal risk of adverse outcomes in PE patients, yet it cannot effectively predict time-to-delivery at the initial diagnosis of PE. To illustrate this, we calculated the maternal risk of adverse outcomes using the fullPIERS formula as reported by von Dadelszen et al[14], used this risk score to estimate the time-to-delivery and compared its performance with our proposed model.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We estimated the probability of adverse outcomes (p) and calculated its concordance index with time to deliver for all PE patients and EOPE subsets, following the original paper. We also plotted the survival curves of high-risk (top 25%), middle-risk (25% - 75%) and low-risk (bottom 25%) groups for all PE and EOPE patients. One limitation is that we do not have chest pain/dyspnoea or SpO2 information collected, so we assume no patients have chest pain or dyspnoea and all patients have 97% SpO2, as instructed by the FullPIERs web calculator https://pre-empt.obgyn.ubc.ca/home-page/past-projects/fullpiers/.

Software

R 4.2.1 and Python were used for all analyses[29,30]. R package "dplyr", "mice" were used in data preparation[31,32]. R package "shiny" and continuumio/anaconda3 Docker image were used to build an interactive web application[28]. Python version 3.9 and R version 4.2.1 are used to run the models in the Docker containers.

RESULT

Cohort characteristics

The discovery cohort consisted of 1,533 PE cases, including 374 EOPE cases collected from the University of Michigan Precision Health and the validation cohort contained 2,172 PE cases (including 547 EOPE) collected from the University of Florida Health System between 2015 and 2022. (**Figure 1**). We employed their EHR data to predict their time-to-delivery from the initial diagnosis of PE. Summaries of the patient characteristics of these cohorts are shown in **Table 1** and **2**.

The baseline prediction model of time to delivery interval among PE patients

PE is a syndrome with well-characterized phenotypes, where hypertension is the most significant clinical symptom. Thus the structured data in the EHR system provide the most useful and straightforward information. From the structured data, we obtained 45 variables including patient demographics, medical history, comorbidities, PE diagnosis time, and severity after data preprocessing (**Supplementary Table**

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2). The resulting model has very decent performance with C-indices of 0.73, 0.72, and 0.71 in the UM cross-validation, UM hold-out testing, and UF validation cohorts, respectively (**Figure 2A**).

To enhance the clinical utilities of the Cox-nnet model, we reduced the number of predictive features following the feature reduction procedure in the Method section. This procedure resulted in 7 significant features, which we used to rebuild the "clinically informative Cox-nnet baseline model". It has C-index scores of 0.73, 0.74, and 0.73 on UM cross-validation, hold-out testing, and UF validation dataset respectively (**Figure 2A**). We stratified patients into 3 groups by the quartiles of predicted time-to-delivery from the reduced model: high-risk (upper quartile), intermediate-risk (interquartile), and low-risk (lower quartile) groups. The survival curves of the time to delivery interval on these three risk groups display significant differences (log-rank p-value < 0.0001) on both the hold-out testing set (**Figure 2B**) and validation set (**Figure 2E**), confirming the strong discriminatory power of the PI score. To enhance the interpretability of the prognosis modeling, we stratified this model using the threshold of 2/7/14 days and predicted the accuracies of delivery using these classifications. The AUROC scores of these classification tasks are 0.85, 0.88, and 0.89 on the testing set **(Figure 2C)** and 0.67, 0.76, and 0.75 on the validation set **(Figure 2F**), respectively.

The seven features in the clinically informative baseline model included those that shorten the time to delivery and extend the time to delivery (**Figure 2D; Table 3**). In descending order of importance scores, the features that shorten the time to delivery are gestational age at diagnosis, sPE, uncomplicated pregestational diabetes mellitus, and parity. Conversely, features extending the time to delivery are PE in a prior pregnancy, increasing maternal age, and comorbid valvular disease. To demonstrate the associations of these important features with time to delivery, we dichotomized patient survival in the hold-out testing set by the median value of each feature (**Supplementary Figure 3**). All features, except maternal age, show significant differences (log-rank p-value < 0.05) between the dichotomized survival groups. We further examined the relationship of the top 3 features (gestational age at diagnosis, sPE, and

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

history of PE in prior pregnancy) with the gestational age at delivery and time to delivery (day) using the UM discovery set in (**Figure 2G-2L).** Later gestational age at diagnosis leads to a later gestational age of delivery (**Figure 2G**), but a shorter time to delivery (**Figure 2J**). sPE is associated with earlier gestational age of delivery (**Figure 2H**) and shorter time to delivery (**Figure 2K**) are diagnosed with sPE. In the deliveries from smaller (<32 weeks) gestational ages, the percentages of patients with PE in prior pregnancies are significantly higher (**Figure 2I)**. However, the percentages of prior PE fluctuate with respect to time to delivery (**Figure 2L**).

Worth noticing, that not all patients diagnosed with PE in 37 weeks or later delivered the babies right away, despite being the least severe cases and can be delivered quickly according to the medical recommendation[12]. Nevertheless, we alternatively built another baseline model with only those patients diagnosed before 37 weeks of gestation. We observed very similar results as the above baseline model using all PE patients, in terms of C-index, the selected top features and their feature scores (**Supplementary Figure 4 A-D**).

The full model of time to delivery among PE patients

We next investigated the contribution to time of delivery from all 60 variables, including the 45 baseline variables above and an additional 15 laboratory testing results and vital signs obtained in the 5-day observation window before the time of diagnosis (**Supplementary Table 2**). The clinical informative model after feature reduction consists of 12 top features (**Table 3**). This model shows significantly ($P < 0.001$, t-test) higher cross-validation accuracy of time to delivery compared to the seven-feature baseline model, with median C-index scores almost as high as 0.80, with 0.78, 0.79, and 0.74 in the cross-validation, testing, and validation datasets respectively. These C-indices are excellent for survival predictions, despite the high heterogeneity of PE and the large patient size which makes it difficult to predict delivery time precisely[18,33]. The Kaplan-Meier curves of the high-, intermediate- and low-risk groups show more significant distinction in testing (**Figure 3B)** and validation set (**Figure 3E**), than the baseline model (**Figure 2B and 2E**). Similarly, we stratified the full model using the threshold of 2/7/14

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

days and predicted the accuracies of delivery using these classifications. The AUROC scores of these classification tasks are 0.88, 0.93, and 0.93 on the testing set and 0.84, 0.89, and 0.90 on the validation set respectively **(Figure 2C, 2F**).

Further examination of the 12 important features in the full model (**Figure 3D, Table 3**) shows good consistency with the 7-feature baseline model (**Figure 2D, Table 3)**. Five out of seven features in the baseline model also exist in the full model with similar importance scores: gestational age at diagnosis, sPE, parity, maternal age, and PE in prior pregnancies. Gestational age at PE diagnosis and sPE continued to be the two most important features in the full model. We also identify new important features from lab tests and vital signs: aspartate aminotransferase (AST) value, the standard deviation of diastolic blood pressure (DBP), the standard deviation of respiratory rate (RR), creatinine value, mean DBP and white blood cell count (**Figure 3D**). Conversely, platelet count is a new feature with a negative importance score, associated with a longer time to delivery. All dichotomized survival plots using median stratification on each of the 12 important features have log-rank p-values smaller than 0.05, confirming their associations with time to delivery in the discovery set (**Supplementary Figure 5)**. We examined the 3 top lab/vital sign features: AST, the standard deviation of DBP, and the standard deviation of RR, on their association with the duration of time between diagnosis and delivery. These values show negative trends with time to delivery, particularly for AST value and the standard deviation of DBP (**Figure 3G-I**). These 3 features are roughly uniformly distributed across delivery gestational ages, except AST which shows slightly higher values in deliveries less than 32 weeks of gestational age (**Supplementary Figure 6**).

Similar to the baseline model earlier using PE patients diagnosed before 37 weeks of gestation, we again alternatively built another full model with the same patients before 37 weeks of gestation. We observed very similar results as the full model using all PE patients, in terms of C-index, the selected top features and their feature scores (**Supplementary Figure 4 E-H**).

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Time to delivery prediction of EOPE patients

Accurate prediction of EOPE patients' time to delivery is crucial, given that delivery of a premature infant has more significant neonatal consequences. Using similar modeling techniques, we trained two additional EOPE-specific Cox-nnet v2 models (baseline vs. full model), using the same features described earlier (**Supplementary Table 2**), on a subset of 374 EOPE patients from the UM discovery cohort. The C-indices for the clinically informative EOPE baseline model are 0.67, 0.68, and 0.63 on the UM cross-validation, hold-out testing, and UF validation sets, respectively (**Figure 4A**). Such significantly lower C-indices for EOPE compared to PE are expected, as EOPE cases are usually difficult to predict prognosis. Still, the time-to-delivery prediction for EOPE is on par or better than the prediction of PE diagnosis using the same set of EHR data[33], demonstrating its potential clinical utility. The KM curves of different predicted survival groups have significant distinctions in both the testing and validation datasets (**Figure 4B** and **4E).** This baseline model consists of the six most important features: gestational age at diagnosis, sPE, PE in a past pregnancy, parity, pulmonary circulatory disorders, and coagulopathies (**Figure 4D; Table 3**). All survival plots, dichotomized using the median stratification on each of the 6 features, have log-rank p-values smaller than 0.05 in the discovery dataset (**Supplementary Figure 7)**. Additionally, the AUROCs of binarized classification on delivery in the next 2/7/14 days range from 0.64-0.82 on the testing set (**Figure 4C**) and 0.52-0.68 on the validation set **(Figure 4F).**

The clinically informative EOPE full model reached much higher accuracy compared to the EOPE baseline model, with median C-indices of 0.74, 0.76, and 0.67 on the cross-validation, testing, and validation sets (**Figure 4G**). The large increases in C-indices are the results of including additional lab tests and blood pressure measurements right around the time of diagnosis of EOPE, confirming their significant clinical values. The 3 risk-stratified groups within the EOPE patient's cohort also showed significant (log-rank p-value<0.001) differences in the hold-out testing set and validation set (**Figure 4H, 4K).** The AUROCs of chance of delivery in the next 2/7/14 days are significantly improved, ranging

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

from 0.82-0.86 on the testing set (**Figure 4I**) and 0.71-0.72 on the validation set **(Figure 4L).** This model contains 13 important features selected from the original 60 features (**Figure 4J; Table 3**). Gestational age at diagnosis continued to be the most important feature. Several other features (eg. PE with severe symptoms, PE in a past pregnancy, parity, and coagulopathy) were of significant importance as well, similar to the EOPE baseline model. Many additional features in the vital signs and lab test categories were also significant, including creatinine value, mean DBP and mean SBP, standard deviation of RR, AST, and platelet counts. Among these 13 features, parity, PE in a prior pregnancy, and higher platelet counts were protective against early delivery **(Figure 4J**).

We created dichotomized survival curves based on creatinine value and platelet count, two new features relative to the EOPE baseline model. Both show strong distinctions between the risk groups (**Figure 4I, 4L**), similar to all other selected features (**Supplementary Figure 7- 8**). These two features also revealed systematic trends in associations with the gestational age at delivery and time from diagnosis to delivery**.** Patients with high creatinine levels were more likely to be delivered within 3 days or less of diagnosis and more likely to deliver preterm (**Figure 4M-4O**). Lower platelet counts were also associated with shorter time to delivery (**Figure 4Q**), even though the platelet levels were not strongly associated with gestational age at delivery among all EOPE patients (**Figure 4R**).

PE time to deliver predictor graphic user interface (GUI)

To disseminate our model publicly, we packaged the pre-trained clinically informative models above into an interactive, user-friendly web application using R shiny[23]. We named this app “PE time to delivery predictor”. The app contains two main panels: the single-patient prediction panel and the group prediction panel (**Supplementary Figure 9**). The single-patient prediction panel calculates the prognosis index (PI) of a single patient if provided the required clinical variables. The PI score describes the patient’s risk of delivery at the time of the diagnosis of PE, relative to the population. The panel also provides the percentile of the PI score among the training data and displays the results in a histogram figure and a

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

table. The group prediction panel calculates the PI and PI percentile of multiple patients simultaneously and also displays them in a table, below the histogram built on the training data. The app is available at http://garmiregroup.org/PE-prognosis-predictor/app

Comparison with previous maternal risk prediction models

Lastly, the previously established maternal risk prediction models (i.e. fullPIERS) cannot effectively predict time-to-delivery at the initial diagnosis of PE directly. We calculated the maternal risk of adverse outcomes using the fullPIERS formula on the UM EHR data. We then used this risk score to estimate the time-to-delivery and compared its performance with our proposed model (see **Methods**). The cross-validation C-index of fullPIERS is 0.50±0.005 on all PE patients and 0.60±0.01 on the EOPE subset (**Supplementary Figure 10A**), much worse than those from our models. So are the survival curves grouped by predicted risk (**Supplementary Figure 10B-C**). Thus the time-to-delivery models are not only different but also irreplaceable by the maternal risk prediction models.

Discussion

PE is a highly heterogenous pregnancy syndrome currently without cure except for delivering the baby and placenta[3,34]. Here we report a new type of survival model to precisely predict the time to delivery as early as the initial diagnosis of PE, subsequent to our recent success in predicting the onset of PE using the same set of EHR data[33]. It helps to save the effort of close monitoring and extensive testing which is conventionally done in resource-rich settings. The simple yet precise models can also be utilized in healthcare systems in resource-limited countries and regions. With such information, clinicians may allocate limited resources in busy antepartum and neonatal ICU beds, or make decisions about the urgency to transfer a patient to a higher level of care in the lack of sufficient resources. As many pregnant women are willing to accept personal risks to improve perceived fetal outcomes, a more concrete model

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

such as the one proposed here will allow them to understand the likely latency and may help them to prepare for delivery emotionally. Many previous studies, such as the fullPIERS and PREP-S models recommended by NICE guidelines, did not predict the precise time of delivery, instead, they fall into very different classification models [16,18,27,28] that aim to predict risks of maternal adverse outcomes (**Supplementary Table 5**). Assisting in deciding delivery timing is not their primary purpose. If they were to be used to predict the time-to-delivery directly, the result would not be satisfactory (**Supplementary Figure 10**). Additionally, the prediction window of proposed models is longer than 48 hours in the fullPIERS model, making them good initial assessment tools.

The proposed models confirmed key factors already highlighted in current PE management, including gestational age at the time of diagnosis, sPE, and the use of creatinine, platelet counts and AST as risk factors in clinical guidelines (**Figure 5**)[12–14,18]. This is not surprising, as less time to delivery is likely associated with patients at higher risk for complications based on clinical assessments. However, this class of models also assigns weights of relative importance, among these key factors, a capacity nonexistent in the current ACOG guidelines[12]. Another novel finding is the identification of parity and PE in prior pregnancies as important predictors for delivery timing in all models tested but not included in current guidelines for PE delivery timing (**Figure 5)**. Most importantly, the models predict the timing of delivery at the initial diagnosis and require no more than readily available information from blood work, medical history, and demographics that are routinely collected in medical centers in the US.

There are several noticeable strengths of this study. The models show consistently high performance in survival prediction and classification tasks, better than previous time-to-delivery prediction models using clinical data or biomarkers[37–39]. Unlike the majority of previous studies that are not validated with external data[35,40,41], our models are validated with an external and independent EHR dataset from UF Health System, despite the noticeable differences between the populations in the two cohorts (**Table 1, 2**). These models also address clinical interpretability by providing importance scores with directionality for

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

each included predictor. Furthermore, the model is designed for accessibility by utilizing fewer than 15 common demographic and disease histories and routinely collected clinical variables in a short observation window. Our approach is much more convenient, as compared to previous studies relying extensively on nonstandard biomarkers such as uterine artery pulsatility index (UtA-PI) or placental growth factor (PLGF)[39–41]. Measurement of these biomarkers is rare in routine prenatal checkups, particularly in lower-income regions, limiting the wide adoption of these biomarker-based models. To maximize the dissemination of the models among clinicians and patients, we have packaged the pre-trained models into a user-friendly shiny application. We aim to embed these models into the EHR system, though it will require additional higher levels of cooperation within the UM Health System. Once integrated, the models will provide clinicians with a fast and accurate assessment of the urgency for delivery at the initial diagnosis of PE.

A few caveats to this study are potentially limiting. Firstly, the ICD coding system lags behind the most recent diagnostic guidelines of PE. However, the codes that are entered into the EHR are based on the clinical assessments of the treating physicians at two academic medical centers, therefore they most likely reflect contemporary diagnostic standards. Also as a retrospective study, the delivery timing can be influenced by clinicians' previous judgment, changes in hospital protocol, communication between patient and provider, intensive care resource availability, and each patient's intentions. However, since these models generally perform well on the external validation set as they achieve C-indices of 0.7 or even close to 0.80, we believe that these subjective factors may not be the main concerns for achieving high prediction power, rather, additional data modalities may help. Prospective investigations of this model's performance in other medical centers would be necessary to confirm the findings. Lastly, our data came from two medical centers with high levels of obstetrics care and therefore testing the model in other settings (eg. other countries and rural regions) will be deemed valuable.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In summary, we have developed the first accurate, deep-learning-based, time-to-delivery prediction models for PE and EOPE patients. The models are disseminated with an easy-to-use web app. Adoption of these models could provide clinicians and patients with valuable management plans to predict and prepare for the best delivery time of each PE pregnancy. Further prospective investigation of the performance of these models is necessary to provide feedback and potential improvement of these models.

Acknowledgment

We thank Anisa Driscoll and Kate Smith from the University of Michigan Precision Health for providing technical support when extracting data used in this study. We acknowledge the University of Florida Integrated Data Repository (IDR) and the UF Health Office of the Chief Data Officer for providing the analytic data set for this project.

Data Sharing Statement

We are unable to publicly share electronic health records data due to its potential to reveal sensitive patient information. However, interested investigators who met the criteria for accessing sensitive data can contact the University of Michigan Precision Health's Research Scientific Facilitators at PHDataHelp@umich.edu (also see https://research.medicine.umich.edu/our-units/data-office-clinical-translational-research/data-access) to inquire about the UM dataset and the necessary steps regarding ethics committee approval and data sharing agreement. Please contact the UF Health Integrated Data Repository (IDR, https://idr.ufhealth.org/) at IRBDataRequest@ahc.ufl.edu to inquire about the UF dataset and the necessary steps regarding ethics committee approval and data sharing agreement.

Code Availability

Codes used for analysis are available at https://github.com/lanagarmire/PE_delivery

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Declaration of Interest

The authors declare no conflict of interest.

Funding

LXG was supported by grants R01 LM012373 and LM012907 awarded by NLM, R01 HD084633 awarded by NICHD. DJL was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K01DK115632) and the University of Florida Clinical and Translational Science Institute (UL1TR001427). XY is supported by NIH/NIGMS Grant T32GM141746. AM is supported by the National Center for Advancing Translational Science (5TL1TR001428).

Author's Contribution

LG conceived this project and supervised the study, after discussing it with ESL. XY conducted data analysis, implemented the Shiny app, and wrote the manuscript. HKB, ADM, KX, and DJL collaborated on validation using the UF cohort. ESL and ADM provided clinical assessments and assistance. DG assisted with Shiny app editing and troubleshooting. All authors have read, revised, and approved the manuscript.

Reference

1. Ives CW, Sinkey R, Rajapreyar I, Tita ATN, Oparil S. Preeclampsia—Pathophysiology and Clinical Presentations. *Journal of the American College of Cardiology*. 2020;76(14):1690-1702. doi:10.1016/j.jacc.2020.08.014
2. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for Preeclampsia: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2017;317(16):1661. doi:10.1001/jama.2017.3439
3. Roberts JM, Rich-Edwards JW, McElrath TF, Garmire L, Myatt L, for the Global Pregnancy Collaboration. Subtypes of Preeclampsia: Recognition and Determining Clinical Usefulness. *Hypertension*. 2021;77(5):1430-1441. doi:10.1161/HYPERTENSIONAHA.120.14781

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4. Sibai BM. Evaluation and management of severe preeclampsia before 34 weeks' gestation. *American Journal of Obstetrics and Gynecology*. 2011;205(3):191-198. doi:10.1016/j.ajog.2011.07.017
5. von Dadelszen P, Magee LA, Roberts JM. Subclassification of Preeclampsia. *Hypertension in Pregnancy*. 2003;22(2):143-148. doi:10.1081/PRG-120021060
6. Phipps EA, Thadhani R, Benzing T, Karumanchi SA. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. *Nat Rev Nephrol*. 2019;15(5):275-289. doi:10.1038/s41581-019-0119-6
7. Amaral LM, Wallace K, Owens M, LaMarca B. Pathophysiology and Current Clinical Management of Preeclampsia. *Curr Hypertens Rep*. 2017;19(8):61. doi:10.1007/s11906-017-0757-7
8. Duffy J, Cairns A, Richards□Doran D, et al. A core outcome set for pre□eclampsia research: an international consensus development study. *BJOG*. 2020;127(12):1516-1526. doi:10.1111/1471-0528.16319
9. Lisonkova S, Joseph KS. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. *American Journal of Obstetrics and Gynecology*. 2013;209(6):544.e1-544.e12. doi:10.1016/j.ajog.2013.08.019
10. Manuck TA, Rice MM, Bailit JL, et al. Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. *American Journal of Obstetrics and Gynecology*. 2016;215(1):103.e1-103.e14. doi:10.1016/j.ajog.2016.01.004
11. Hollegaard B, Lykke JA, Boomsma JJ. Time from pre-eclampsia diagnosis to delivery affects future health prospects of children. *Evolution, Medicine, and Public Health*. 2017;2017(1):53-66. doi:10.1093/emph/eox004
12. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstetrics & Gynecology*. 2020;135(6):e237-e260. doi:10.1097/AOG.0000000000003891
13. Visintin C, Mugglestone MA, Almerie MQ, et al. Management of hypertensive disorders during pregnancy: summary of NICE guidance. *BMJ*. 2010;341(aug25 3):c2207-c2207. doi:10.1136/bmj.c2207
14. 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
15. Ukah UV, Payne B, Hutcheon JA, et al. Assessment of the fullPIERS Risk Prediction Model in Women With Early-Onset Preeclampsia. *Hypertension*. 2018;71(4):659-665. doi:10.1161/HYPERTENSIONAHA.117.10318
16. Ukah UV, Payne B, Karjalainen H, et al. Temporal and external validation of the fullPIERS model for the prediction of adverse maternal outcomes in women with pre-eclampsia. *Pregnancy Hypertension*. 2019;15:42-50. doi:10.1016/j.preghy.2018.01.004
17. 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. Lawn JE, ed. *PLoS Med*. 2014;11(1):e1001589. doi:10.1371/journal.pmed.1001589
18. for the PREP Collaborative Network, Thangaratinam S, Allotey J, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(1):68. doi:10.1186/s12916-017-0827-3
19. Schmidt LJ, Rieger O, Neznansky M, et al. A machine-learning-based algorithm improves prediction of preeclampsia-associated adverse outcomes. *Am J Obstet Gynecol*. 2022;227(1):77.e1-77.e30. doi:10.1016/j.ajog.2022.01.026
20. Wang D, Jing Z, He K, Garmire LX. Cox-nnet v2.0: improved neural-network-based survival prediction extended to large-scale EMR data. Schwartz R, ed. *Bioinformatics*. 2021;37(17):2772-2774. doi:10.1093/bioinformatics/btab046
21. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. Markowetz F, ed. *PLoS Comput Biol*. 2018;14(4):e1006076. doi:10.1371/journal.pcbi.1006076
22. Zhan Z, Jing Z, He B, et al. Two-stage Cox-nnet: biologically interpretable neural-network model for

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

prognosis prediction and its application in liver cancer survival using histopathology and transcriptomic data. *NAR Genomics and Bioinformatics*. 2021;3(1):lqab015. doi:10.1093/nargab/lqab015

23. Zawistowski M, Fritsche LG, Pandit A, et al. The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genomics*. 2023;3(2):100257. doi:10.1016/j.xgen.2023.100257
24. ACOG Practice Bulletin No. 203: Chronic Hypertension in Pregnancy. *Obstetrics & Gynecology*. 2019;133(1):e26-e50. doi:10.1097/AOG.0000000000003020
25. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340-1347. doi:10.1093/bioinformatics/btq134
26. Harrell FE. Evaluating the Yield of Medical Tests. *JAMA*. 1982;247(18):2543. doi:10.1001/jama.1982.03320430047030
27. Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society Series A (General)*. 1972;135(2):185. doi:10.2307/2344317
28. Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, Jonathan McPherson. shiny: Web Application Framework for R. Published online 2020. https://CRAN.R-project.org/package=shiny
29. R Core Team. R: A language and environment for statistical computing. Published online 2021. https://www.R-project.org/
30. Rossum G van, Drake FL. The Python language reference. Published online 2010.
31. Wickham H, François R, Lionel Henry, Kirill Müller. dplyr: A Grammar of Data Manipulation. Published online 2020. https://CRAN.R-project.org/package=dplyr
32. Buuren S van, Groothuis-Oudshoorn K. **mice**: Multivariate Imputation by Chained Equations in *R*. *J Stat Soft*. 2011;45(3). doi:10.18637/jss.v045.i03
33. Ballard HK, Yang X, Mahadevan A, Lemas DJ, Garmire LX. Building and validating 5-feature models to predict preeclampsia onset time from electronic health record data. Published online March 24, 2023. doi:10.1101/2023.03.23.23287655
34. Benny PA, Alakwaa FM, Schlueter RJ, Lassiter CB, Garmire LX. A review of omics approaches to study preeclampsia. *Placenta*. 2020;92:17-27. doi:10.1016/j.placenta.2020.01.008
35. Li S, Wang Z, Vieira LA, et al. Improving preeclampsia risk prediction by modeling pregnancy trajectories from routinely collected electronic medical record data. *npj Digit Med*. 2022;5(1):68. doi:10.1038/s41746-022-00612-x
36. Jhee JH, Lee S, Park Y, et al. Prediction model development of late-onset preeclampsia using machine learning-based methods. Spradley FT, ed. *PLoS ONE*. 2019;14(8):e0221202. doi:10.1371/journal.pone.0221202
37. Duhig KE, Seed PT, Placzek A, et al. Prognostic indicators of severe disease in late preterm pre-eclampsia to guide decision making on timing of delivery: The PEACOCK study. *Pregnancy Hypertension*. 2021;24:90-95. doi:10.1016/j.preghy.2021.02.012
38. Reeder HT, Haneuse S, Modest AM, Hacker MR, Sudhof LS, Papatheodorou SI. A novel approach to joint prediction of preeclampsia and delivery timing using semicompeting risks. *American Journal of Obstetrics and Gynecology*. Published online August 2022:S0002937822006883. doi:10.1016/j.ajog.2022.08.045
39. Lim S, Li W, Kemper J, Nguyen A, Mol BW, Reddy M. Biomarkers and the Prediction of Adverse Outcomes in Preeclampsia: A Systematic Review and Meta-analysis. *Obstetrics & Gynecology*. 2021;137(1):72-81. doi:10.1097/AOG.0000000000004149
40. Binder J, Palmrich P, Pateisky P, et al. The Prognostic Value of Angiogenic Markers in Twin Pregnancies to Predict Delivery Due to Maternal Complications of Preeclampsia. *Hypertension*. 2020;76(1):176-183. doi:10.1161/HYPERTENSIONAHA.120.14957
41. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. *American Journal of Obstetrics and Gynecology*. 2020;223(1):12-23.e7. doi:10.1016/j.ajog.2019.11.1247

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tables and legends

Table 1: Basic Patient Characteristics for this study

Variable Name	**University of Michigan PE** ***Discovery Cohort (n = 1533)***	**University of Florida PE** ***Validation Cohort (n = 2172)***
Maternal Age, mean (SD)	*30.19(5.74)*	*28.63(6.46)*
Parity, mean (SD)	*0.72(1.15)*	*0.73(1.55)*
Number of fetuses, mean (SD)	*1.07(0.27)*	*1.04(0.21)*
Gestational Age at Diagnosis (days), mean (SD)	*249.96(27.44)*	*256.72(26.83)*
Time to Delivery (days), mean (SD)	*7.02(15.39)*	*5.70(13.28)*
Race, N (%)		
African American	*0.18*	*0.37*
American Indian or Alaska Native	*0.00*	*0.00*
Asian	*0.07*	*0.02*
Caucasian	*0.75*	*0.49*
Native Hawaiian and Other Pacific Islander	*0.00*	*0.00*
Unknown or Other	*0.00*	*0.12*
Ethnicity, N (%)		
Hispanic	*0.06*	*0.11*
Non-Hispanic	*0.94*	*0.87*
Unknown	*0.00*	*0.01*
Smoking Status, N (%)		
Current Smoker	*0.06*	*0.13*
Former Smoker	*0.24*	*NA*
Never Smoker	*0.70*	*0.87*
Illegal Drug Use Status, N(%)		
Yes	*0.10*	*0.17*
No	*0.90*	*0.83*
History of PE, N (%)		
Yes	*0.11*	*0.12*
No	*0.89*	*0.88*
sPE, N (%)		
Yes	*0.47*	*0.35*
No	*0.53*	*0.65*

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Basic Characteristics of EOPE patients in this study

Variable Name	University of Michigan EOPE Subset (n = 374)	University of Florida EOPE Subset (n=547)
Maternal Age, mean (SD)	30.70(5.72)	29.48(6.65)
Parity, mean (SD)	0.98(1.26)	1.11(1.55)
Number of fetuses, mean (SD)	1.12(0.34)	1.09(0.31)
Gestational Age at Diagnosis, mean (SD), day	211.10(23.79)	217.75(21.43)
Time to Delivery, mean (SD), day	19.12(26.44)	16.28 (22.12)
Race, N (%)		
African American	26.74%	43.33%
American Indian or Alaska Native	0.80%	0.18%
Asian	5.08%	1.10%
Caucasian	66.58%	45.16%
Native Hawaiian and Other Pacific Islander	0.80%	0.00%
Unknown or Other	0.00%	10.23%
Ethnicity, N (%)		
Hispanic	3.21%	9.14%
Non-Hispanic	96.79%	89.58%
Unknown	0.00%	1.28%
Smoking Status, N (%)		
Current Smoker	7.22%	17.00%
Former Smoker	22.99%	NA
Never Smoker	69.79%	NA
Illegal Drug Use Status, N(%)		
Yes	11.76%	14.08%
No	88.24%	85.92%
History of PE, N (%)		
Yes	18.72%	15.54%
No	81.28%	84.46%
sPE, N (%)		
Yes	54.28%	35.10%
No	45.72%	64.90%

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3: Features and their permutation importance score in each reduced model

Baseline Model for all PE		Full model for all PE		Baseline Model for EOPE		Full Model for EOPE	
Name	**Importance Score**	**Name**	**Importance Score**	**Name**	**Importance Score**	**Name**	**Importance Score**
Gestational Age at Diagnosis	*323.55*	*Gestational Age at Diagnosis*	*390.29*	*Gestational Age at Diagnosis*	*29.88*	*Gestational Age at Diagnosis*	*26.36*
sPE	*79.59*	*sPE*	*38.71*	*sPE*	*12.37*	*Creatinine*	*10.52*
PE in Prior Pregnancies	*7.02*	*DBP Standard Deviation*	*30.22*	*PE in Prior Pregnancies*	*6.69*	*Mean DBP*	*8.85*
Maternal Age	*2.75*	*AST*	*26.56*	*Pulmonary Circulatory Disorders*	*6.38*	*Mean SBP*	*7.21*
Uncomplicated Diabetes	*1.35*	*Mean DBP*	*13.64*	*Parity*	*4.78*	*AST*	*7.1*
Valvular Disease	*1.11*	*RR Standard Deviation*	*12.37*	*Coagulopathies*	*1.17*	*Pulmonary Circulatory Disorders*	*5.96*
Parity	*1.04*	*Parity*	*11.45*			*Parity*	*4.74*
		Creatinine	*11.16*			*RR Standard Deviation*	*4.52*
		Platelet Count	*9.67*			*sPE*	*2.83*
		White Blood Cell Count	*6.08*			*Number of Fetuses*	*2.24*
		Age	*3.84*			*Platelet Count*	*1.93*
		PE in Prior Pregnancies	*3.38*			*PE in Prior Pregnancy*	*1.34*
						Coagulopathies	*0.69*

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure Legends

Figure 1: Experimental design and cohort overview. A) Experiment Design Workflow: The discovery cohort was obtained from the University of Michigan Health System and a validation cohort of similar size and time was obtained from the University of Florida Health System. We constructed 4 predictive models: baseline and full models for all PE patients and baseline and full models for EOPE patients. The input variables in baseline models include patients' demographics, lifestyle, comorbidities and medical history. The full models include additional lab tests and vital signs from within 5 days of PE diagnosis, in addition to the variables in the baseline models. We trained the Cox-nnet prognosis prediction model using 80% training from the discovery cohort, tested it on 20% hold-out data from the discovery cohort, and validated it using the validation cohort. We then built clinically informative models by reducing Cox-nnet features based on both their importance scores and significance levels. The models are examined by the importance scores of top features and stratified survival curves based on patient survival risks. We disseminated the feature-reduced, clinically informative models into a user-friendly web application for healthcare professionals to use. Created with BioRender.com.

Figure 2: PE Baseline model results, interpretation, and evaluation. A: The bar plots of C-indices from the original Cox-nnet models (red) and feature-reduced clinically informative model (green), on the UM cross-validation and hold-out testing set and UF validation set. B, E: The survival curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted PI from the reduced baseline model in A on B: hold-out testing data and E: validation data. C, F: ROC curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced baseline model on C: hold-out testing data and F: validation data. D: The ln-transformed permutation importance scores of features in the feature-reduced baseline model. A positive sign indicates that a higher value in feature is associated with a shorter time to delivery and a negative sign means an extension of time to delivery. G-L: The distribution of diagnosis gestational age, sPE rate and PE in prior pregnancy rate, in associations with delivery gestational week (G-I) and time (days) to delivery (J-L).

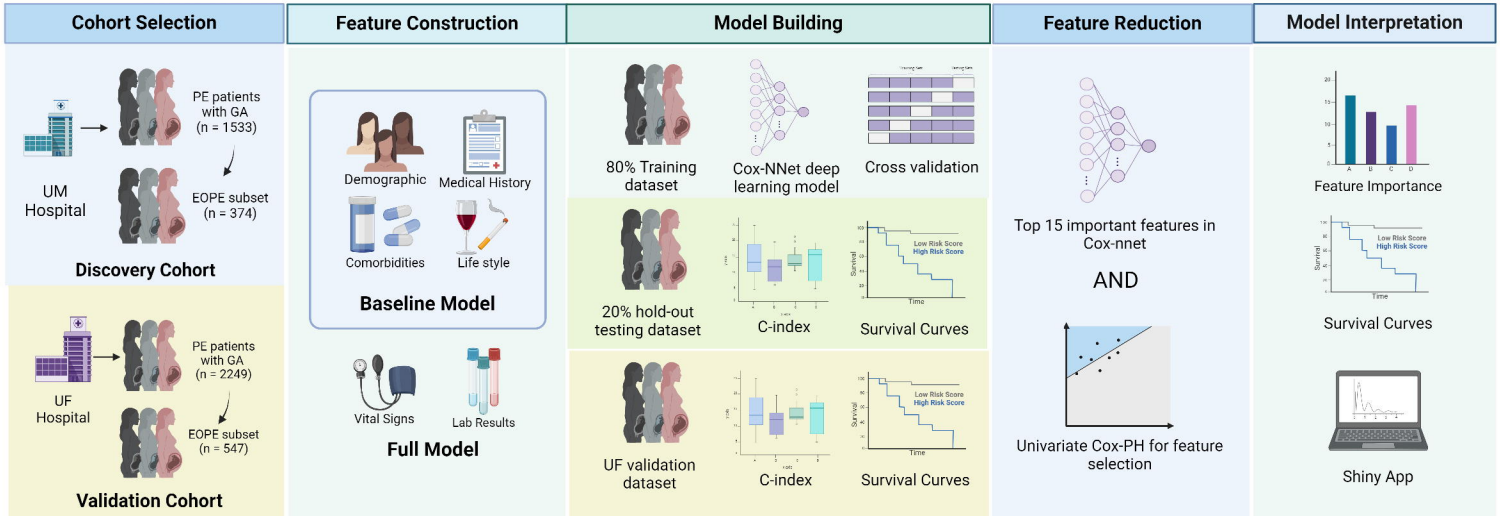
Figure 3: PE Full model results, interpretation and evaluation. A: The bar plots of C-indices from the original models (red) and feature-reduced clinical informative model (green), on the UM training cross-validation and hold-out testing set and UF validation set. B, E: The survival curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted PI from the reduced full model in A. B: hold-out testing data, E: validation data. C, F: ROC curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced full model (A) on C: hold-out testing data and F: validation data. D: The ln-transformed permutation importance scores of features in the feature-reduced baseline model. A positive sign indicates that a higher value in the feature

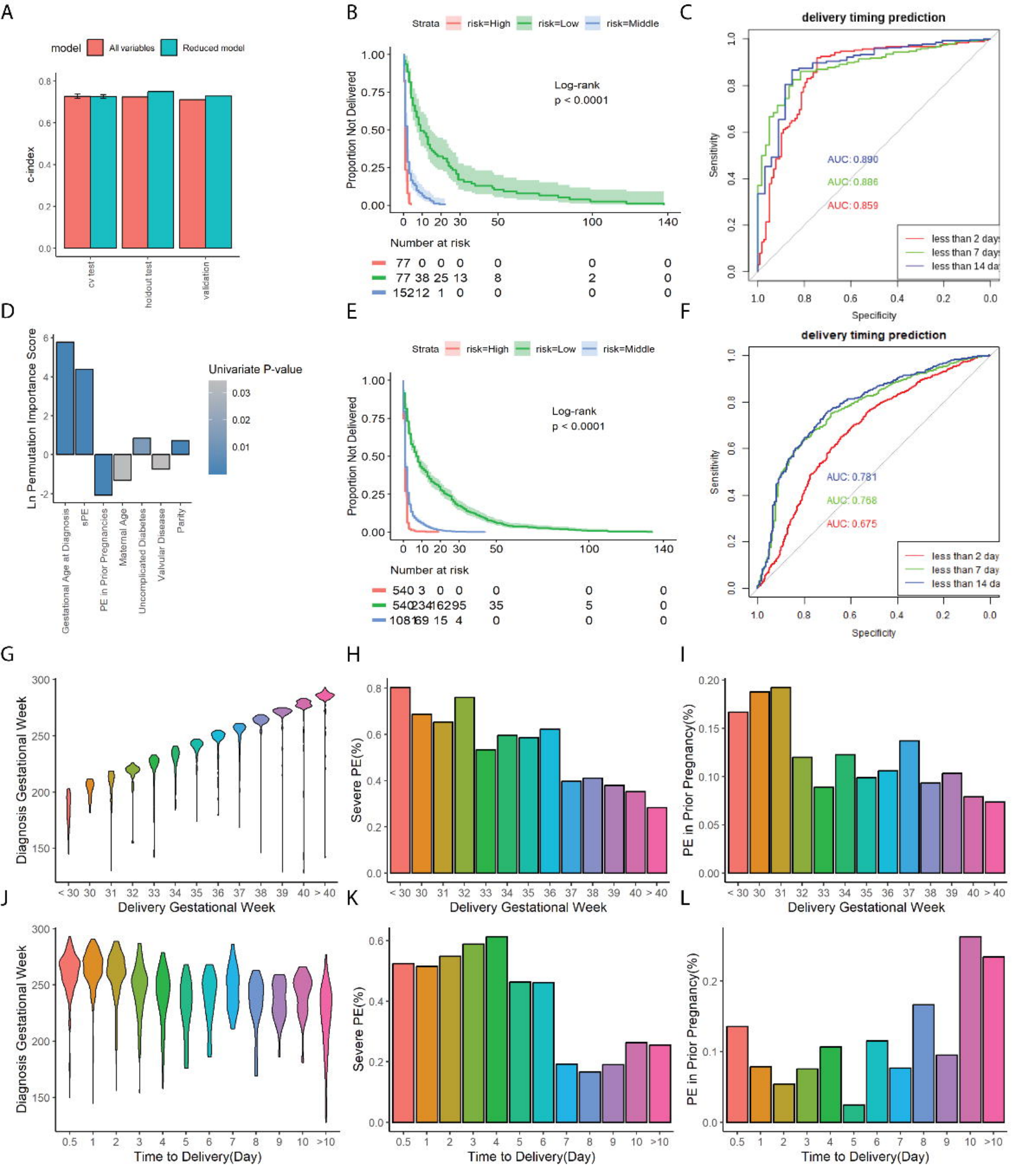
medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273366; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

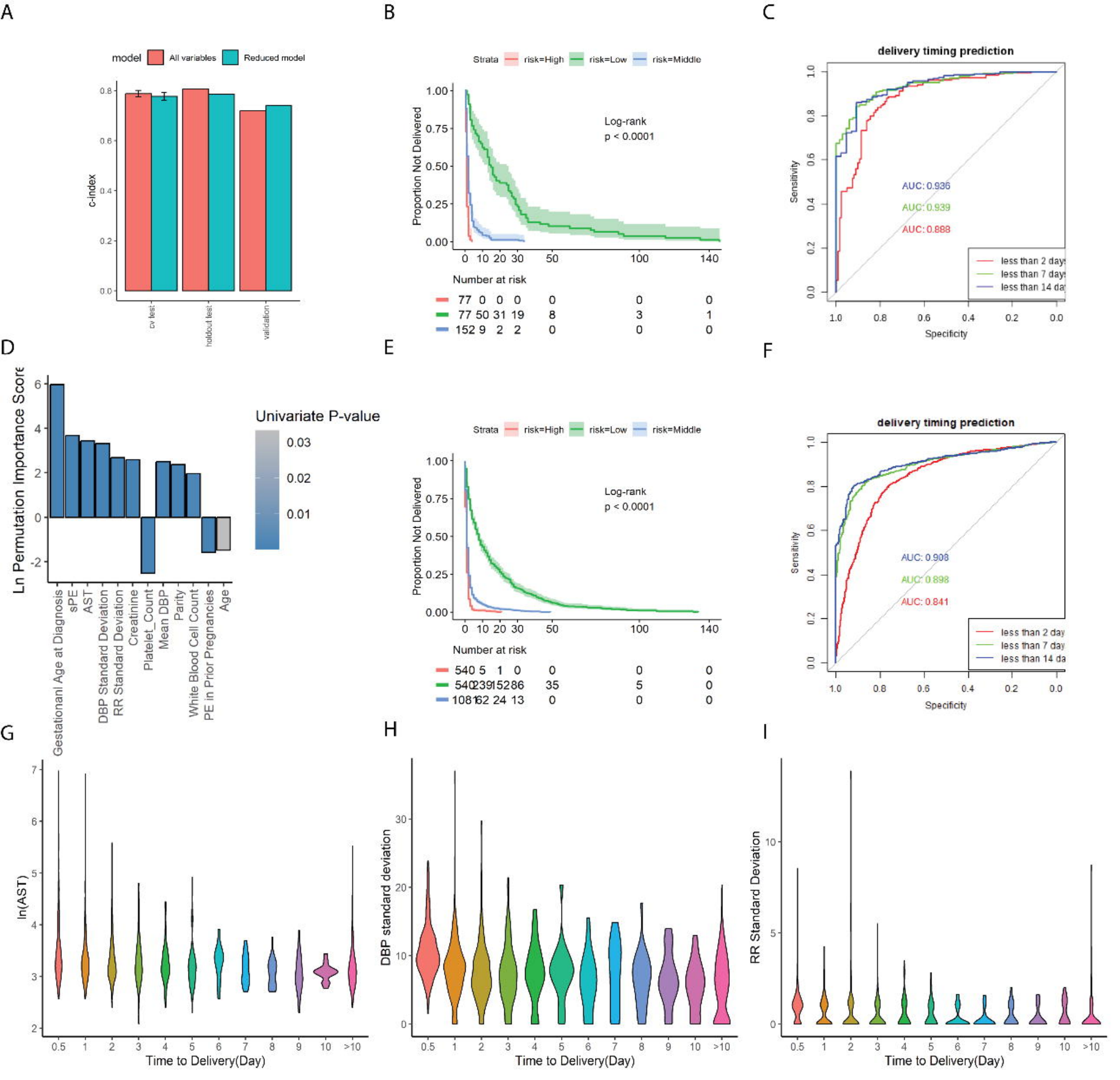
is associated with a shorter time to delivery and a negative sign means an extension of time to delivery. G-I: The distribution of aspartate aminotransferase (AST) values, the standard deviation of diastolic blood pressure (DBP) and the standard deviation of respiratory rate (RR), in association with time (days) to delivery.

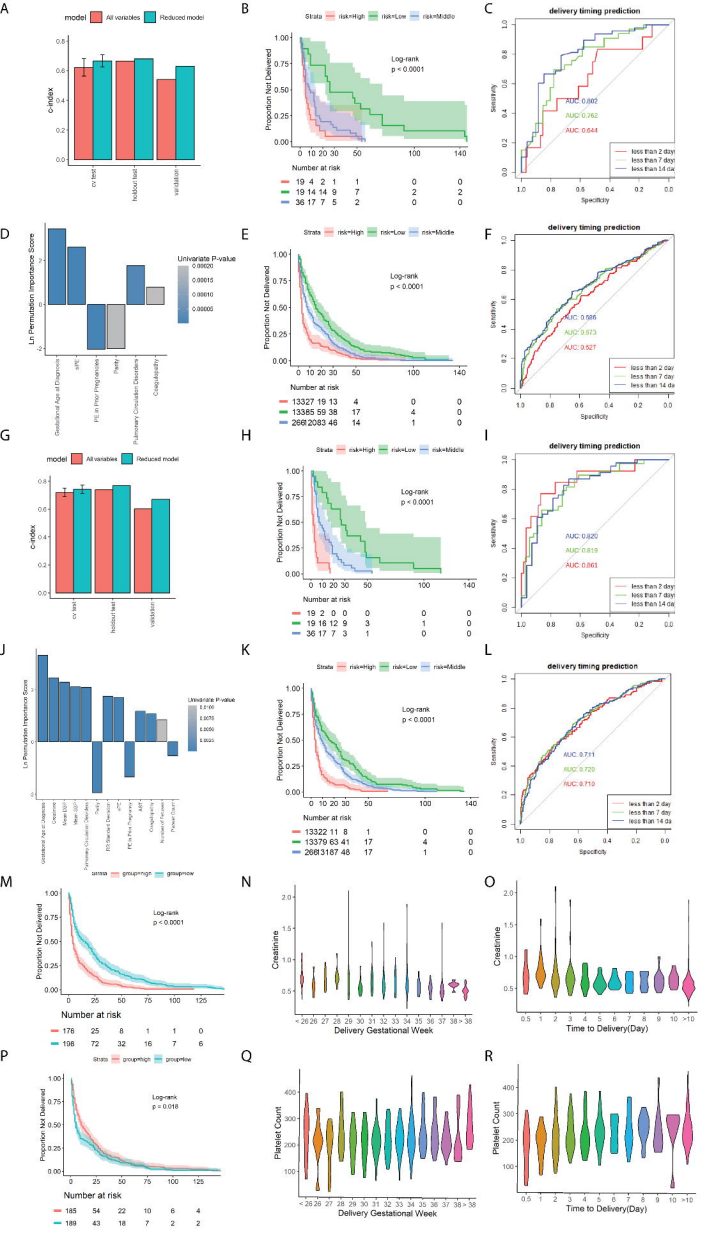
Figure 4: Results, interpretation and evaluation of baseline and full models on the EOPE patient subset. A: The bar plots of C-indices from the original Cox-nnet EOPE baseline model (red) and feature-reduced clinically informative model (green) on the cross-validation and testing set and validation set. B-C: The survival curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted PI from the reduced EOPE baseline model in A. B, E: hold-out testing data, E: validation data. C, F: ROC curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced EOPE baseline model (A) on C: hold-out testing data and F: validation data. D: The ln-transformed permutation importance score of features in the EOPE full model. G: The bar plots of C-indices from the original Cox-nnet EOPE full model (red) and its feature-reduced clinically informative model (green) on the cross-validation and testing set and validation set. H, K: The survival curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted PI from the reduced full model in E. H: hold-out testing data, K: validation data. I, L: ROC curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced EOPE full model (G) on I: hold-out testing data and L: validation data. J: The ln-transformed permutation importance scores of features in the EOPE full model. M-O: Analysis of creatinine values among the EOPE patients in the discovery cohort. M: The dichotomized survival curves by the median value of creatinine. N, O: Distributions of creatinine values by delivery gestational week (N) and time to delivery (O). P-R: Analysis of platelet counts among the EOPE patients in the discovery cohort. P: The dichotomized survival curves by the median value of platelet counts. Q-R: Distributions of creatinine values by delivery gestational week (Q) and time to delivery (R).

Figure 5: Comparison of important features among the four feature-reduced clinically informative models A: The bubble plot of important features from PE baseline, EOPE baseline, PE full, and EOPE full models using reduced top important features. The size of the bubbles represents the permutation importance score of each feature. Color represents the sign of features in the time to delivery prediction: a positive sign indicates that a higher value in the feature is associated with a shorter time to delivery and a negative sign means an extension of time to delivery. B: Venn diagram of the important features from the four models shown in A.


Cohort Selection
PE patients with GA (n = 1533)
UM Hospital
EOPE subset (n = 374)
Discovery Cohort
PE patients with GA (n = 2249)
UF Hospital
EOPE subset (n = 547)
Validation Cohort
Feature Construction
Demographic
Medical History
Comorbidities
Life style
Baseline Model
Vital Signs
Lab Results
Full Model
Model Building
80% Training dataset
Cox-NNet deep learning model
Cross validation
20% hold-out testing dataset
C-index
Survival Curves
Low Risk Score
High Risk Score
Time
UF validation dataset
C-index
Survival Curves
Feature Reduction
Top 15 important features in Cox-nnet
AND
Univariate Cox-PH for feature selection
Model Interpretation
Feature Importance
Survival Curves
Shiny App


A
model
All variables
Reduced model
c-index
cv test
holdout test
validation
B
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
77 0 0 0 0 0 0
77 38 25 13 8 2 0
15212 1 0 0 0 0
C
delivery timing prediction
Sensitivity
Specificity
AUC: 0.890
AUC: 0.886
AUC: 0.859
less than 2 days
less than 7 days
less than 14 da
D
Ln Permutation Importance Score
Univariate P-value
Gestational Age at Diagnosis
sPE
PE in Prior Pregnancies
Maternal Age
Uncomplicated Diabetes
Valvular Disease
Parity
E
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
540 3 0 0 0 0 0
540234162 95 35 5 0
108169 15 4 0 0 0
F
delivery timing prediction
Sensitivity
Specificity
AUC: 0.781
AUC: 0.768
AUC: 0.675
less than 2 day
less than 7 day
less than 14 da
G
Diagnosis Gestational Week
Delivery Gestational Week
H
Severe PE(%)
Delivery Gestational Week
I
PE in Prior Pregnancy(%)
Delivery Gestational Week
J
Diagnosis Gestational Week
Time to Delivery(Day)
K
Severe PE(%)
Time to Delivery(Day)
L
PE in Prior Pregnancy(%)
Time to Delivery(Day)

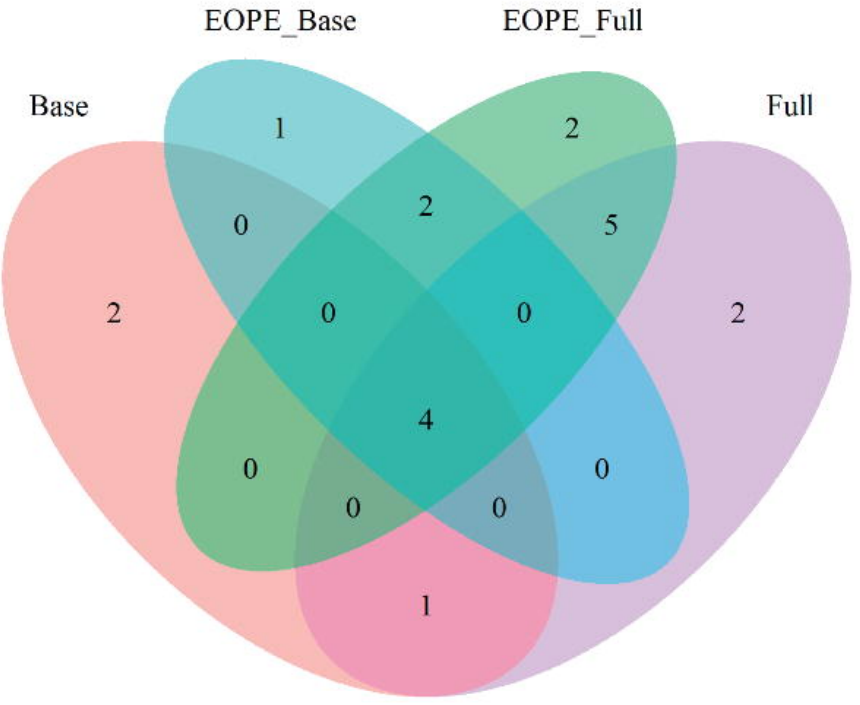

A
model
All variables
Reduced model
c-index
cv test
holdout test
validation
B
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
77 0 0 0 0 0 0
77 50 31 19 8 3 1
152 9 2 2 0 0 0
C
delivery timing prediction
Sensitivity
Specificity
AUC: 0.936
AUC: 0.939
AUC: 0.888
less than 2 days
less than 7 days
less than 14 da
D
Ln Permutation Importance Score
Univariate P-value
Gestationanl Age at Diagnosis
sPE
AST
DBP Standard Deviation
RR Standard Deviation
Creatinine
Platelet_Count
Mean DBP
Parity
White Blood Cell Count
PE in Prior Pregnancies
Age
E
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
540 5 1 0 0 0 0
540 239 152 86 35 5 0
108 162 24 13 0 0 0
F
delivery timing prediction
Sensitivity
Specificity
AUC: 0.908
AUC: 0.898
AUC: 0.841
less than 2 day
less than 7 day
less than 14 da
G
ln(AST)
Time to Delivery(Day)
H
DBP standard deviation
Time to Delivery(Day)
I
RR Standard Deviation
Time to Delivery(Day)


A
model
All variables
Reduced model
c-index
cv test
holdout test
validation
B
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
19 4 2 1 1 0 0
19 14 14 9 7 2 2
36 17 7 5 2 0 0
C
delivery timing prediction
Sensitivity
Specificity
AUC: 0.802
AUC: 0.762
AUC: 0.644
less than 2 days
less than 7 days
less than 14 da
D
Ln Permutation Importance Score
Univariate P-value
0.00020
0.00015
0.00010
0.00005
Gestational Age at Diagnosis
aPC
PE in Prior Pregnancies
Parity
Pulmonary Circulation Disorders
Coagulopathy
E
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
13327 19 13 4 0 0
13385 59 38 17 4 0
266 2083 46 14 1 0
F
delivery timing prediction
Sensitivity
Specificity
AUC: 0.586
AUC: 0.573
AUC: 0.527
less than 2 day
less than 7 day
less than 14 da
G
model
All variables
Reduced model
c-index
cv test
holdout test
validation
H
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
19 2 0 0 0 0 0
19 16 12 9 3 1 0
36 17 7 3 1 0 0
I
delivery timing prediction
Sensitivity
Specificity
AUC: 0.820
AUC: 0.819
AUC: 0.861
less than 2 days
less than 7 days
less than 14 da
J
Ln Permutation Importance Score
Univariate P-value
Gestational Age at Diagnosis
Coagulopathy
Pulmonary Circulation Disorders
Parity
PE in Prior Pregnancies
aPC
Coagulopathy
Number of Fetuses
Patient Count
K
Strata
risk=High
risk=Low
risk=Middle
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
13322 11 8 1 0 0
13379 63 41 17 4 0
266 3187 48 17 1 0
L
delivery timing prediction
Sensitivity
Specificity
AUC: 0.711
AUC: 0.720
AUC: 0.710
less than 2 day
less than 7 day
less than 14 da
M
Strata
group=high
group=low
Proportion Not Delivered
Log-rank
p < 0.0001
Number at risk
176 25 8 1 1 0
198 72 32 16 7 6
N
Creatinine
Delivery Gestational Week
O
Creatinine
Time to Delivery(Day)
P
Strata
group=high
group=low
Proportion Not Delivered
Log-rank
p = 0.018
Number at risk
185 54 22 10 6 4
189 43 18 7 2 2
Q
Platelet Count
Delivery Gestational Week
R
Platelet Count
Time to Delivery(Day)

A

Gestational Age at Diagnosis
sPE
PE in Prior Pregnancies
Parity
AST
DBP Standard Deviation
RR Standard Deviation
Creatinine
Platelet Count
Mean DBP
White Blood Cell Count
Maternal Age
Uncomplicated Diabetes
Valvular Disease
Pulmonary Circulation Disorders
Coagulopathy
Mean SBP
Number of Fetuses
name
base
eope_base
full
eope_full
model
Importance
100
200
300
sign
Negative
Positive

B


EOPE_Base
EOPE_Full
Base
Full
1
2
0
2
5
2
0
0
2
4
0
0
0
0
1