

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

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

6  
7 1 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor,  
8 Michigan, United States of America

9 2 Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville,  
10 Florida, United States of America

11 3 Department of Obstetrics & Gynecology, University of Florida College of Medicine, Gainesville,  
12 Florida

13 4 Department of Physiology and Aging, University of Florida, Gainesville, FL, United States

14 5 Center for Research in Perinatal Outcomes, University of Florida, Gainesville, FL, United States

15 6 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor,  
16 Michigan, United States of America

17 7 Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, Michigan, United  
18 States of America

19

20

21

22

23 \*: Corresponding author:

24 Lana Garmire

25 [lgarmire@umich.edu](mailto:lgarmire@umich.edu)

26 Room 3366, Building 520, NCRC,

27 1600 Huron Parkway, Ann Arbor, MI 48105

28 Office phone: (734) 615-0514

29

30

31 word count: 5067

32

33

34 **Abbreviations:**

35 PE: preeclampsia

36 sPE: Preeclampsia with severe features

37 EOPE: early-onset preeclampsia

38 LOPE: late-onset preeclampsia

39 EHR: electronic health record

40 SBP: systolic blood pressure

41 DBP: diastolic blood pressure

42 RR: respiratory rate

43 HELLP: hemolysis, elevated liver enzymes, low platelet count

44 AST: aspartate transaminase

45 PI: prognosis score

46 UM: University of Michigan

47 UF: University of Florida

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

49 MAP: mean arterial pressure

50 UtA-PI: uterine artery pulsatility index

51 PLGF: placental growth factor

52 ACOG: American College of Obstetricians and Gynecologist

53 AUROC: Area Under the Receiver-Operator curve

54

55

56

57

58

59

60 **Abstract**

61 **Background**

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

68 **Method**

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

77

78 **Results**

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

84

85 **Conclusions**

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

89

90

## 91 **Introduction**

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

107

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

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

128  
129 Toward this goal, we developed and externally validated the first deep learning model to predict patient  
130 delivery time after the initial diagnosis of PE using electronic health records (EHR) data. We utilized the  
131 state-of-the-art deep learning-based prognosis prediction model, Cox-nnet (version 2), which we  
132 previously developed<sup>20-22</sup>. Cox-nnet methods had previously consistently shown excellent predictive  
133 performances under a variety of conditions, including on EHR data<sup>20</sup>. Our objectives were: (1) to predict  
134 the time to delivery at the first diagnosis of PE for all PE patients and an EOPE sub-cohort, by  
135 constructing and validating deep-learning models utilizing EHR data; and (2) to assess the quantitative

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

138

## 139 **Methods**

### 140 **Data Source**

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

156

### 157 **Fully connected Cox-nnet neural network models**

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

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

162

### 163 **EHR Feature Engineering**

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

172

173 Pregnancy characteristics included parity, number of fetuses, gestational age, PE severity at initial  
174 diagnosis, and history of preterm birth, c-section, abruption, etc. Fetal development includes poor fetal  
175 growth according to the associated ICD code(O36.59). Other comorbidities were grouped into 29  
176 categories using the Elixhauser Comorbidity Index<sup>1</sup>. The observational window for lab results and vital  
177 signs was 5 days before the day of the initial PE diagnosis. Only the first results of repeated lab tests were  
178 used to avoid intervention/treatment effects. Summary statistics of systolic blood pressure (SBP),  
179 diastolic blood pressure (DBP), and respiratory rate (RR) measures were included (max, min, mean,  
180 standard deviation), as done in previous work<sup>2</sup>. We removed features with high missing proportions (over  
181 20%) and sparse features with fewer than 10 non-zero values. Highly correlated variables were identified  
182 using the variance inflation factor (VIF) and removed one at a time until all features had a VIF below 3 to  
183 avoid multicollinearity. The remaining missing values were imputed using the PMM algorithm from R  
184 package “mice”. All numerical features were scaled by dividing their root mean square. Numeric features

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

187

### 188 **Reduced feature representation from the Cox-nnet models**

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

198

### 199 **Model evaluation**

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



211

## 212 **Interactive Web Application for Easy Model Validation**

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

221

## 222 **External Validation using UF data**

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

228

## 229 **Estimate time-to-deliver using maternal risk of adverse outcomes calculated from the fullPIERS** 230 **model**

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

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

242

### 243 **Software**

244 R 4.2.1 and Python were used for all analyses<sup>29,30</sup>. R package “dplyr”, “mice” were used in data  
245 preparation<sup>31,32</sup>. R package “shiny” and continuumio/anaconda3 Docker image were used to build an  
246 interactive web application<sup>28</sup>. Python version 3.9 and R version 4.2.1 are used to run the models in the  
247 Docker containers.

248

## 249 **RESULT**

### 250 **Cohort characteristics**

251 The discovery cohort consisted of 1,533 PE cases, including 374 EOPE cases collected from the  
252 University of Michigan Precision Health and the validation cohort contained 2,172 PE cases (including  
253 547 EOPE) collected from the University of Florida Health System between 2015 and 2022. (**Figure 1**).  
254 We employed their EHR data to predict their time-to-delivery from the initial diagnosis of PE. Summaries  
255 of the patient characteristics of these cohorts are shown in **Table 1** and **2**.

256

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

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

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

264  
265 To enhance the clinical utilities of the Cox-nnet model, we reduced the number of predictive features  
266 following the feature reduction procedure in the Method section. This procedure resulted in 7 significant  
267 features, which we used to rebuild the “clinically informative Cox-nnet baseline model”. It has C-index  
268 scores of 0.73, 0.74, and 0.73 on UM cross-validation, hold-out testing, and UF validation dataset  
269 respectively (**Figure 2A**). We stratified patients into 3 groups by the quartiles of predicted time-to-  
270 delivery from the reduced model: high-risk (upper quartile), intermediate-risk (interquartile), and low-risk  
271 (lower quartile) groups. The survival curves of the time to delivery interval on these three risk groups  
272 display significant differences (log-rank p-value < 0.0001) on both the hold-out testing set (**Figure 2B**)  
273 and validation set (**Figure 2E**), confirming the strong discriminatory power of the PI score. To enhance  
274 the interpretability of the prognosis modeling, we stratified this model using the threshold of 2/7/14 days  
275 and predicted the accuracies of delivery using these classifications. The AUROC scores of these  
276 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  
277 validation set (**Figure 2F**), respectively.

278  
279 The seven features in the clinically informative baseline model included those that shorten the time to  
280 delivery and extend the time to delivery (**Figure 2D; Table 3**). In descending order of importance scores,  
281 the features that shorten the time to delivery are gestational age at diagnosis, sPE, uncomplicated  
282 pregestational diabetes mellitus, and parity. Conversely, features extending the time to delivery are PE in  
283 a prior pregnancy, increasing maternal age, and comorbid valvular disease. To demonstrate the  
284 associations of these important features with time to delivery, we dichotomized patient survival in the  
285 hold-out testing set by the median value of each feature (**Supplementary Figure 3**). All features, except  
286 maternal age, show significant differences (log-rank p-value < 0.05) between the dichotomized survival  
287 groups. We further examined the relationship of the top 3 features (gestational age at diagnosis, sPE, and

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

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

301

### 302 **The full model of time to delivery among PE patients**

303 We next investigated the contribution to time of delivery from all 60 variables, including the 45 baseline  
304 variables above and an additional 15 laboratory testing results and vital signs obtained in the 5-day  
305 observation window before the time of diagnosis (**Supplementary Table 2**). The clinical informative  
306 model after feature reduction consists of 12 top features (**Table 3**). This model shows significantly ( $P <$   
307  $0.001$ , t-test) higher cross-validation accuracy of time to delivery compared to the seven-feature baseline  
308 model, with median C-index scores almost as high as 0.80, with 0.78, 0.79, and 0.74 in the cross-  
309 validation, testing, and validation datasets respectively. These C-indices are excellent for survival  
310 predictions, despite the high heterogeneity of PE and the large patient size which makes it difficult to  
311 predict delivery time precisely<sup>18,33</sup>. The Kaplan-Meier curves of the high-, intermediate- and low-risk  
312 groups show more significant distinction in testing (**Figure 3B**) and validation set (**Figure 3E**), than the  
313 baseline model (**Figure 2B and 2E**). Similarly, we stratified the full model using the threshold of 2/7/14

314 days and predicted the accuracies of delivery using these classifications. The AUROC scores of these  
315 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  
316 respectively (**Figure 2C, 2F**).

317  
318 Further examination of the 12 important features in the full model (**Figure 3D, Table 3**) shows good  
319 consistency with the 7-feature baseline model (**Figure 2D, Table 3**). Five out of seven features in the  
320 baseline model also exist in the full model with similar importance scores: gestational age at diagnosis,  
321 sPE, parity, maternal age, and PE in prior pregnancies. Gestational age at PE diagnosis and sPE continued  
322 to be the two most important features in the full model. We also identify new important features from lab  
323 tests and vital signs: aspartate aminotransferase (AST) value, the standard deviation of diastolic blood  
324 pressure (DBP), the standard deviation of respiratory rate (RR), creatinine value, mean DBP and white  
325 blood cell count (**Figure 3D**). Conversely, platelet count is a new feature with a negative importance  
326 score, associated with a longer time to delivery. All dichotomized survival plots using median  
327 stratification on each of the 12 important features have log-rank p-values smaller than 0.05, confirming  
328 their associations with time to delivery in the discovery set (**Supplementary Figure 5**). We examined the  
329 3 top lab/vital sign features: AST, the standard deviation of DBP, and the standard deviation of RR, on  
330 their association with the duration of time between diagnosis and delivery. These values show negative  
331 trends with time to delivery, particularly for AST value and the standard deviation of DBP (**Figure 3G-I**).  
332 These 3 features are roughly uniformly distributed across delivery gestational ages, except AST which  
333 shows slightly higher values in deliveries less than 32 weeks of gestational age (**Supplementary Figure**  
334 **6**).

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

339

## 340 **Time to delivery prediction of EOPE patients**

341 Accurate prediction of EOPE patients' time to delivery is crucial, given that delivery of a premature infant  
342 has more significant neonatal consequences. Using similar modeling techniques, we trained two additional  
343 EOPE-specific Cox-nnet v2 models (baseline vs. full model), using the same features described earlier  
344 (**Supplementary Table 2**), on a subset of 374 EOPE patients from the UM discovery cohort.

345 The C-indices for the clinically informative EOPE baseline model are 0.67, 0.68, and 0.63 on the UM  
346 cross-validation, hold-out testing, and UF validation sets, respectively (**Figure 4A**). Such significantly  
347 lower C-indices for EOPE compared to PE are expected, as EOPE cases are usually difficult to predict  
348 prognosis. Still, the time-to-delivery prediction for EOPE is on par or better than the prediction of PE  
349 diagnosis using the same set of EHR data<sup>33</sup>, demonstrating its potential clinical utility. The KM curves of  
350 different predicted survival groups have significant distinctions in both the testing and validation datasets  
351 (**Figure 4B** and **4E**). This baseline model consists of the six most important features: gestational age at  
352 diagnosis, sPE, PE in a past pregnancy, parity, pulmonary circulatory disorders, and coagulopathies  
353 (**Figure 4D; Table 3**). All survival plots, dichotomized using the median stratification on each of the 6  
354 features, have log-rank p-values smaller than 0.05 in the discovery dataset (**Supplementary Figure 7**).  
355 Additionally, the AUROCs of binarized classification on delivery in the next 2/7/14 days range from  
356 0.64-0.82 on the testing set (**Figure 4C**) and 0.52-0.68 on the validation set (**Figure 4F**).

357  
358 The clinically informative EOPE full model reached much higher accuracy compared to the EOPE  
359 baseline model, with median C-indices of 0.74, 0.76, and 0.67 on the cross-validation, testing, and  
360 validation sets (**Figure 4G**). The large increases in C-indices are the results of including additional lab  
361 tests and blood pressure measurements right around the time of diagnosis of EOPE, confirming their  
362 significant clinical values. The 3 risk-stratified groups within the EOPE patient's cohort also showed  
363 significant (log-rank p-value<0.001) differences in the hold-out testing set and validation set (**Figure 4H,**  
364 **4K**). The AUROCs of chance of delivery in the next 2/7/14 days are significantly improved, ranging

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

373  
374 We created dichotomized survival curves based on creatinine value and platelet count, two new features  
375 relative to the EOPE baseline model. Both show strong distinctions between the risk groups (**Figure 4I,**  
376 **4L**), similar to all other selected features (**Supplementary Figure 7- 8**). These two features also revealed  
377 systematic trends in associations with the gestational age at delivery and time from diagnosis to delivery.  
378 Patients with high creatinine levels were more likely to be delivered within 3 days or less of diagnosis and  
379 more likely to deliver preterm (**Figure 4M-4O**). Lower platelet counts were also associated with shorter  
380 time to delivery (**Figure 4Q**), even though the platelet levels were not strongly associated with gestational  
381 age at delivery among all EOPE patients (**Figure 4R**).

382

### 383 **PE time to deliver predictor graphic user interface (GUI)**

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

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

394

### 395 **Comparison with previous maternal risk prediction models**

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

404

405

### 406 **Discussion**

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



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

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

434  
435 There are several noticeable strengths of this study. The models show consistently high performance in  
436 survival prediction and classification tasks, better than previous time-to-delivery prediction models using  
437 clinical data or biomarkers<sup>37-39</sup>. Unlike the majority of previous studies that are not validated with  
438 external data<sup>35,40,41</sup>, our models are validated with an external and independent EHR dataset from UF  
439 Health System, despite the noticeable differences between the populations in the two cohorts (**Table 1, 2**).  
440 These models also address clinical interpretability by providing importance scores with directionality for

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

452  
453 A few caveats to this study are potentially limiting. Firstly, the ICD coding system lags behind the most  
454 recent diagnostic guidelines of PE. However, the codes that are entered into the EHR are based on the  
455 clinical assessments of the treating physicians at two academic medical centers, therefore they most likely  
456 reflect contemporary diagnostic standards. Also as a retrospective study, the delivery timing can be  
457 influenced by clinicians' previous judgment, changes in hospital protocol, communication between  
458 patient and provider, intensive care resource availability, and each patient's intentions. However, since  
459 these models generally perform well on the external validation set as they achieve C-indices of 0.7 or  
460 even close to 0.80, we believe that these subjective factors may not be the main concerns for achieving  
461 high prediction power, rather, additional data modalities may help. Prospective investigations of this  
462 model's performance in other medical centers would be necessary to confirm the findings. Lastly, our  
463 data came from two medical centers with high levels of obstetrics care and therefore testing the model in other  
464 settings (eg. other countries and rural regions) will be deemed valuable.

465

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

472

### 473 **Acknowledgment**

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

478

### 479 **Data Sharing Statement**

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

488

### 489 **Code Availability**

490 Codes used for analysis are available at [https://github.com/lanagarmire/PE\\_delivery](https://github.com/lanagarmire/PE_delivery)

491

492 **Declaration of Interest**

493 The authors declare no conflict of interest.

494

495 **Funding**

496 LXG was supported by grants R01 LM012373 and LM012907 awarded by NLM, R01 HD084633

497 awarded by NICHD. DJL was supported by the National Institute of Diabetes and Digestive and Kidney

498 Diseases (K01DK115632) and the University of Florida Clinical and Translational Science Institute

499 (UL1TR001427). XY is supported by NIH/NIGMS Grant T32GM141746. AM is supported by the

500 National Center for Advancing Translational Science (5TL1TR001428).

501

502 **Author's Contribution**

503 LG conceived this project and supervised the study, after discussing it with ESL. XY conducted data

504 analysis, implemented the Shiny app, and wrote the manuscript. HKB, ADM, KX, and DJL collaborated

505 on validation using the UF cohort. ESL and ADM provided clinical assessments and assistance. DG

506 assisted with Shiny app editing and troubleshooting. All authors have read, revised, and approved the

507 manuscript.

508

509 **Reference**

510 1. Ives CW, Sinkey R, Rajapreyar I, Tita ATN, Oparil S. Preeclampsia—Pathophysiology and Clinical  
511 Presentations. *Journal of the American College of Cardiology*. 2020;76(14):1690-1702.

512 doi:10.1016/j.jacc.2020.08.014

513 2. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for  
514 Preeclampsia: US Preventive Services Task Force Recommendation Statement. *JAMA*.

515 2017;317(16):1661. doi:10.1001/jama.2017.3439

516 3. Roberts JM, Rich-Edwards JW, McElrath TF, Garmire L, Myatt L, for the Global Pregnancy  
517 Collaboration. Subtypes of Preeclampsia: Recognition and Determining Clinical Usefulness.

518 *Hypertension*. 2021;77(5):1430-1441. doi:10.1161/HYPERTENSIONAHA.120.14781

- 519 4. Sibai BM. Evaluation and management of severe preeclampsia before 34 weeks' gestation. *American*  
520 *Journal of Obstetrics and Gynecology*. 2011;205(3):191-198. doi:10.1016/j.ajog.2011.07.017
- 521 5. von Dadelszen P, Magee LA, Roberts JM. Subclassification of Preeclampsia. *Hypertension in*  
522 *Pregnancy*. 2003;22(2):143-148. doi:10.1081/PRG-120021060
- 523 6. Phipps EA, Thadhani R, Benzinger T, Karumanchi SA. Pre-eclampsia: pathogenesis, novel diagnostics  
524 and therapies. *Nat Rev Nephrol*. 2019;15(5):275-289. doi:10.1038/s41581-019-0119-6
- 525 7. Amaral LM, Wallace K, Owens M, LaMarca B. Pathophysiology and Current Clinical Management  
526 of Preeclampsia. *Curr Hypertens Rep*. 2017;19(8):61. doi:10.1007/s11906-017-0757-7
- 527 8. Duffy J, Cairns A, Richards D, Doran D, et al. A core outcome set for pre-eclampsia research: an  
528 international consensus development study. *BJOG*. 2020;127(12):1516-1526. doi:10.1111/1471-  
529 0528.16319
- 530 9. Lisonkova S, Joseph KS. Incidence of preeclampsia: risk factors and outcomes associated with early-  
531 versus late-onset disease. *American Journal of Obstetrics and Gynecology*. 2013;209(6):544.e1-  
532 544.e12. doi:10.1016/j.ajog.2013.08.019
- 533 10. Manuck TA, Rice MM, Bailit JL, et al. Preterm neonatal morbidity and mortality by gestational age:  
534 a contemporary cohort. *American Journal of Obstetrics and Gynecology*. 2016;215(1):103.e1-  
535 103.e14. doi:10.1016/j.ajog.2016.01.004
- 536 11. Hollegaard B, Lykke JA, Boomsma JJ. Time from pre-eclampsia diagnosis to delivery affects future  
537 health prospects of children. *Evolution, Medicine, and Public Health*. 2017;2017(1):53-66.  
538 doi:10.1093/emph/eox004
- 539 12. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstetrics &*  
540 *Gynecology*. 2020;135(6):e237-e260. doi:10.1097/AOG.0000000000003891
- 541 13. Visintin C, Mugglestone MA, Almerie MQ, et al. Management of hypertensive disorders during  
542 pregnancy: summary of NICE guidance. *BMJ*. 2010;341(aug25 3):c2207-c2207.  
543 doi:10.1136/bmj.c2207
- 544 14. Von Dadelszen P, Payne B, Li J, et al. Prediction of adverse maternal outcomes in pre-eclampsia:  
545 development and validation of the fullPIERS model. *The Lancet*. 2011;377(9761):219-227.  
546 doi:10.1016/S0140-6736(10)61351-7
- 547 15. Ukah UV, Payne B, Hutcheon JA, et al. Assessment of the fullPIERS Risk Prediction Model in  
548 Women With Early-Onset Preeclampsia. *Hypertension*. 2018;71(4):659-665.  
549 doi:10.1161/HYPERTENSIONAHA.117.10318
- 550 16. Ukah UV, Payne B, Karjalainen H, et al. Temporal and external validation of the fullPIERS model  
551 for the prediction of adverse maternal outcomes in women with pre-eclampsia. *Pregnancy*  
552 *Hypertension*. 2019;15:42-50. doi:10.1016/j.preghy.2018.01.004
- 553 17. Payne BA, Hutcheon JA, Ansermino JM, et al. A Risk Prediction Model for the Assessment and  
554 Triage of Women with Hypertensive Disorders of Pregnancy in Low-Resourced Settings: The  
555 miniPIERS (Pre-eclampsia Integrated Estimate of RiSk) Multi-country Prospective Cohort Study.  
556 Lawn JE, ed. *PLoS Med*. 2014;11(1):e1001589. doi:10.1371/journal.pmed.1001589
- 557 18. for the PREP Collaborative Network, Thangaratnam S, Allotey J, et al. Prediction of complications  
558 in early-onset pre-eclampsia (PREP): development and external multinational validation of  
559 prognostic models. *BMC Med*. 2017;15(1):68. doi:10.1186/s12916-017-0827-3
- 560 19. Schmidt LJ, Rieger O, Neznansky M, et al. A machine-learning-based algorithm improves prediction  
561 of preeclampsia-associated adverse outcomes. *Am J Obstet Gynecol*. 2022;227(1):77.e1-77.e30.  
562 doi:10.1016/j.ajog.2022.01.026
- 563 20. Wang D, Jing Z, He K, Garmire LX. Cox-nnet v2.0: improved neural-network-based survival  
564 prediction extended to large-scale EMR data. Schwartz R, ed. *Bioinformatics*. 2021;37(17):2772-  
565 2774. doi:10.1093/bioinformatics/btab046
- 566 21. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction  
567 of high-throughput omics data. Markowetz F, ed. *PLoS Comput Biol*. 2018;14(4):e1006076.  
568 doi:10.1371/journal.pcbi.1006076
- 569 22. Zhan Z, Jing Z, He B, et al. Two-stage Cox-nnet: biologically interpretable neural-network model for

- 570 prognosis prediction and its application in liver cancer survival using histopathology and  
571 transcriptomic data. *NAR Genomics and Bioinformatics*. 2021;3(1):lqab015.  
572 doi:10.1093/nargab/lqab015
- 573 23. Zawistowski M, Fritsche LG, Pandit A, et al. The Michigan Genomics Initiative: A biobank linking  
574 genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genomics*.  
575 2023;3(2):100257. doi:10.1016/j.xgen.2023.100257
- 576 24. ACOG Practice Bulletin No. 203: Chronic Hypertension in Pregnancy. *Obstetrics & Gynecology*.  
577 2019;133(1):e26-e50. doi:10.1097/AOG.0000000000003020
- 578 25. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance  
579 measure. *Bioinformatics*. 2010;26(10):1340-1347. doi:10.1093/bioinformatics/btq134
- 580 26. Harrell FE. Evaluating the Yield of Medical Tests. *JAMA*. 1982;247(18):2543.  
581 doi:10.1001/jama.1982.03320430047030
- 582 27. Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal*  
583 *Statistical Society Series A (General)*. 1972;135(2):185. doi:10.2307/2344317
- 584 28. Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, Jonathan McPherson. shiny: Web Application  
585 Framework for R. Published online 2020. <https://CRAN.R-project.org/package=shiny>
- 586 29. R Core Team. R: A language and environment for statistical computing. Published online 2021.  
587 <https://www.R-project.org/>
- 588 30. Rossum G van, Drake FL. The Python language reference. Published online 2010.
- 589 31. Wickham H, François R, Lionel Henry, Kirill Müller. dplyr: A Grammar of Data Manipulation.  
590 Published online 2020. <https://CRAN.R-project.org/package=dplyr>
- 591 32. Buuren S van, Groothuis-Oudshoorn K. **mice**  $\square$ : Multivariate Imputation by Chained Equations in R.  
592 *J Stat Soft*. 2011;45(3). doi:10.18637/jss.v045.i03
- 593 33. Ballard HK, Yang X, Mahadevan A, Lemas DJ, Garmire LX. Building and validating 5-feature  
594 models to predict preeclampsia onset time from electronic health record data. Published online March  
595 24, 2023. doi:10.1101/2023.03.23.23287655
- 596 34. Benny PA, Alakwaa FM, Schlueter RJ, Lassiter CB, Garmire LX. A review of omics approaches to  
597 study preeclampsia. *Placenta*. 2020;92:17-27. doi:10.1016/j.placenta.2020.01.008
- 598 35. Li S, Wang Z, Vieira LA, et al. Improving preeclampsia risk prediction by modeling pregnancy  
599 trajectories from routinely collected electronic medical record data. *npj Digit Med*. 2022;5(1):68.  
600 doi:10.1038/s41746-022-00612-x
- 601 36. Jhee JH, Lee S, Park Y, et al. Prediction model development of late-onset preeclampsia using  
602 machine learning-based methods. Spradley FT, ed. *PLoS ONE*. 2019;14(8):e0221202.  
603 doi:10.1371/journal.pone.0221202
- 604 37. Duhig KE, Seed PT, Placzek A, et al. Prognostic indicators of severe disease in late preterm pre-  
605 eclampsia to guide decision making on timing of delivery: The PEACOCK study. *Pregnancy*  
606 *Hypertension*. 2021;24:90-95. doi:10.1016/j.preghy.2021.02.012
- 607 38. Reeder HT, Haneuse S, Modest AM, Hacker MR, Sudhof LS, Papatheodorou SI. A novel approach to  
608 joint prediction of preeclampsia and delivery timing using semicompeting risks. *American Journal of*  
609 *Obstetrics and Gynecology*. Published online August 2022:S0002937822006883.  
610 doi:10.1016/j.ajog.2022.08.045
- 611 39. Lim S, Li W, Kemper J, Nguyen A, Mol BW, Reddy M. Biomarkers and the Prediction of Adverse  
612 Outcomes in Preeclampsia: A Systematic Review and Meta-analysis. *Obstetrics & Gynecology*.  
613 2021;137(1):72-81. doi:10.1097/AOG.0000000000004149
- 614 40. Binder J, Palmrich P, Pateisky P, et al. The Prognostic Value of Angiogenic Markers in Twin  
615 Pregnancies to Predict Delivery Due to Maternal Complications of Preeclampsia. *Hypertension*.  
616 2020;76(1):176-183. doi:10.1161/HYPERTENSIONAHA.120.14957
- 617 41. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia.  
618 *American Journal of Obstetrics and Gynecology*. 2020;223(1):12-23.e7.  
619 doi:10.1016/j.ajog.2019.11.1247

620

621

622

623

624

625 **Tables and legends**

626 **Table 1: Basic Patient Characteristics for this study**

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

627

628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642

**Table 2: Basic Characteristics of EOPE patients in this study**

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

643  
644  
645  
646  
647



648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661

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

662

663  
664  
665  
666  
667  
668  
669  
670  
671  
672

## 673 **Figure Legends**

674  
675

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

689  
690 **Figure 2: PE Baseline model results, interpretation, and evaluation. A:** The bar plots of C-indices  
691 from the original Cox-nnet models (red) and feature-reduced clinically informative model (green), on the  
692 UM cross-validation and hold-out testing set and UF validation set. B, E: The survival curves of high-risk  
693 (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted  
694 PI from the reduced baseline model in A on B: hold-out testing data and E: validation data. C, F: ROC  
695 curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced baseline  
696 model on C: hold-out testing data and F: validation data. D: The ln-transformed permutation importance  
697 scores of features in the feature-reduced baseline model. A positive sign indicates that a higher value in  
698 feature is associated with a shorter time to delivery and a negative sign means an extension of time to  
699 delivery. G-L: The distribution of diagnosis gestational age, sPE rate and PE in prior pregnancy rate, in  
700 associations with delivery gestational week (G-I) and time (days) to delivery (J-L).

701  
702

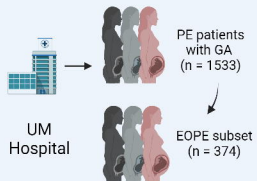
703 **Figure 3: PE Full model results, interpretation and evaluation. A:** The bar plots of C-indices from the  
704 original models (red) and feature-reduced clinical informative model (green), on the UM training cross-  
705 validation and hold-out testing set and UF validation set. B, E: The survival curves of high-risk (top  
706 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%), categorized by predicted PI  
707 from the reduced full model in A. B: hold-out testing data, E: validation data. C, F: ROC curves of  
708 prediction delivery time within 2 days, 7 days and 14 days using results from reduced full model (A) on  
709 C: hold-out testing data and F: validation data. D: The ln-transformed permutation importance scores of  
710 features in the feature-reduced baseline model. A positive sign indicates that a higher value in the feature

711 is associated with a shorter time to delivery and a negative sign means an extension of time to delivery.  
712 G-I: The distribution of aspartate aminotransferase (AST) values, the standard deviation of diastolic blood  
713 pressure (DBP) and the standard deviation of respiratory rate (RR), in association with time (days) to  
714 delivery.  
715

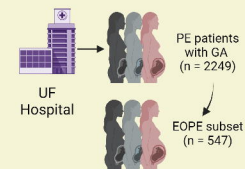
716 **Figure 4: Results, interpretation and evaluation of baseline and full models on the EOPE patient**  
717 **subset.** A: The bar plots of C-indices from the original Cox-nnet EOPE baseline model (red) and feature-  
718 reduced clinically informative model (green) on the cross-validation and testing set and validation set. B-  
719 C: The survival curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups  
720 (bottom 25%), categorized by predicted PI from the reduced EOPE baseline model in A. B, E: hold-out  
721 testing data, E: validation data. C, F: ROC curves of prediction delivery time within 2 days, 7 days and 14  
722 days using results from reduced EOPE baseline model (A) on C: hold-out testing data and F: validation  
723 data. D: The ln-transformed permutation importance score of features in the EOPE full model. G: The bar  
724 plots of C-indices from the original Cox-nnet EOPE full model (red) and its feature-reduced clinically  
725 informative model (green) on the cross-validation and testing set and validation set. H, K: The survival  
726 curves of high-risk (top 25%), intermediate-risk (middle 50%) and low-risk groups (bottom 25%),  
727 categorized by predicted PI from the reduced full model in E. H: hold-out testing data, K: validation data.  
728 I, L: ROC curves of prediction delivery time within 2 days, 7 days and 14 days using results from reduced  
729 EOPE full model (G) on I: hold-out testing data and L: validation data. J: The ln-transformed permutation  
730 importance scores of features in the EOPE full model. M-O: Analysis of creatinine values among the  
731 EOPE patients in the discovery cohort. M: The dichotomized survival curves by the median value of  
732 creatinine. N, O: Distributions of creatinine values by delivery gestational week (N) and time to delivery  
733 (O). P-R: Analysis of platelet counts among the EOPE patients in the discovery cohort. P: The  
734 dichotomized survival curves by the median value of platelet counts. Q-R: Distributions of creatinine  
735 values by delivery gestational week (Q) and time to delivery (R).  
736  
737

738 **Figure 5: Comparison of important features among the four feature-reduced clinically informative**  
739 **models** A: The bubble plot of important features from PE baseline, EOPE baseline, PE full, and EOPE  
740 full models using reduced top important features. The size of the bubbles represents the permutation  
741 importance score of each feature. Color represents the sign of features in the time to delivery prediction: a  
742 positive sign indicates that a higher value in the feature is associated with a shorter time to delivery and a  
743 negative sign means an extension of time to delivery. B: Venn diagram of the important features from the  
744 four models shown in A.  
745

## Cohort Selection

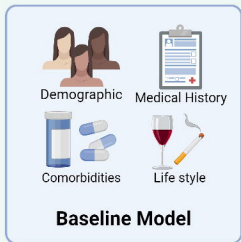


### Discovery Cohort



### Validation Cohort

## Feature Construction

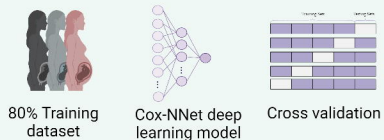


### Baseline Model



### Full Model

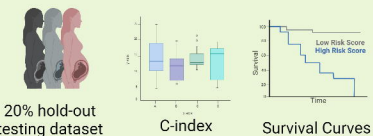
## Model Building



80% Training dataset

Cox-NNet deep learning model

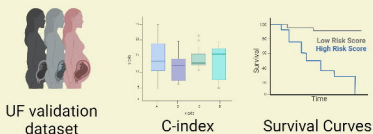
Cross validation



20% hold-out testing dataset

C-index

Survival Curves

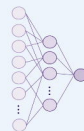


UF validation dataset

C-index

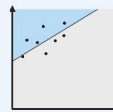
Survival Curves

## Feature Reduction



Top 15 important features in Cox-nnet

AND

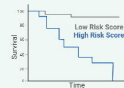


Univariate Cox-PH for feature selection

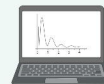
## Model Interpretation



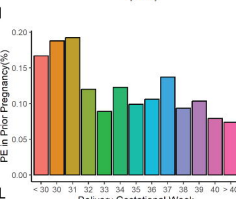
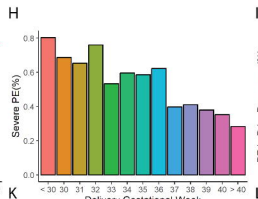
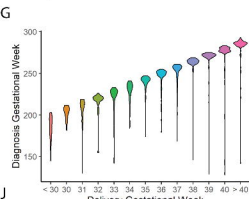
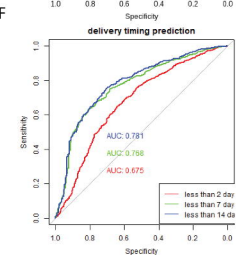
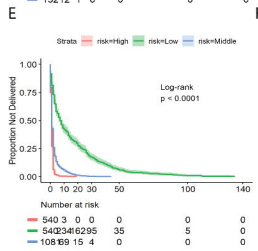
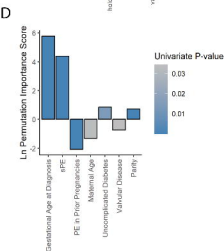
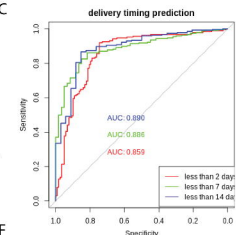
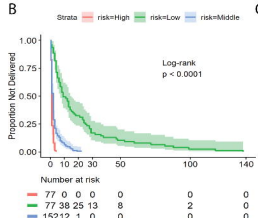
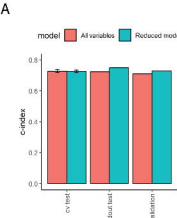
Feature Importance

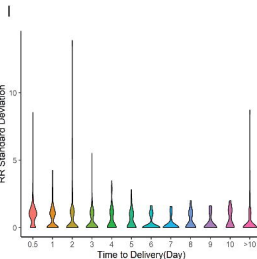
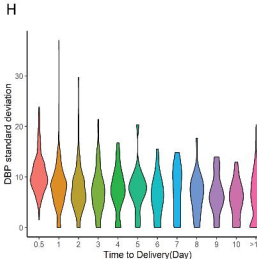
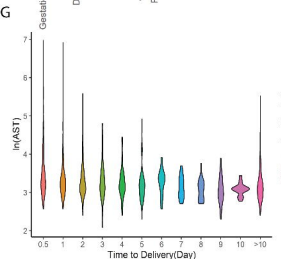
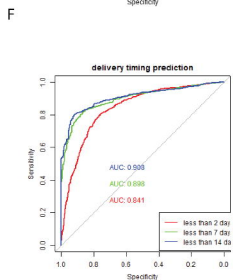
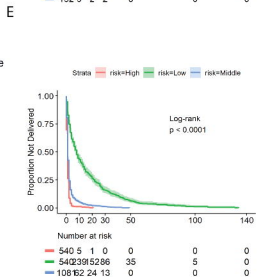
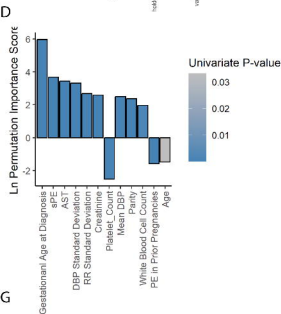
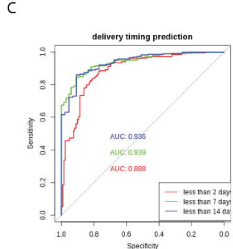
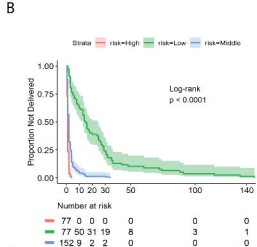
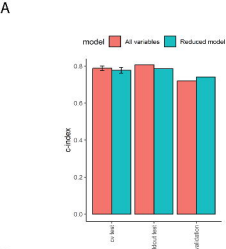


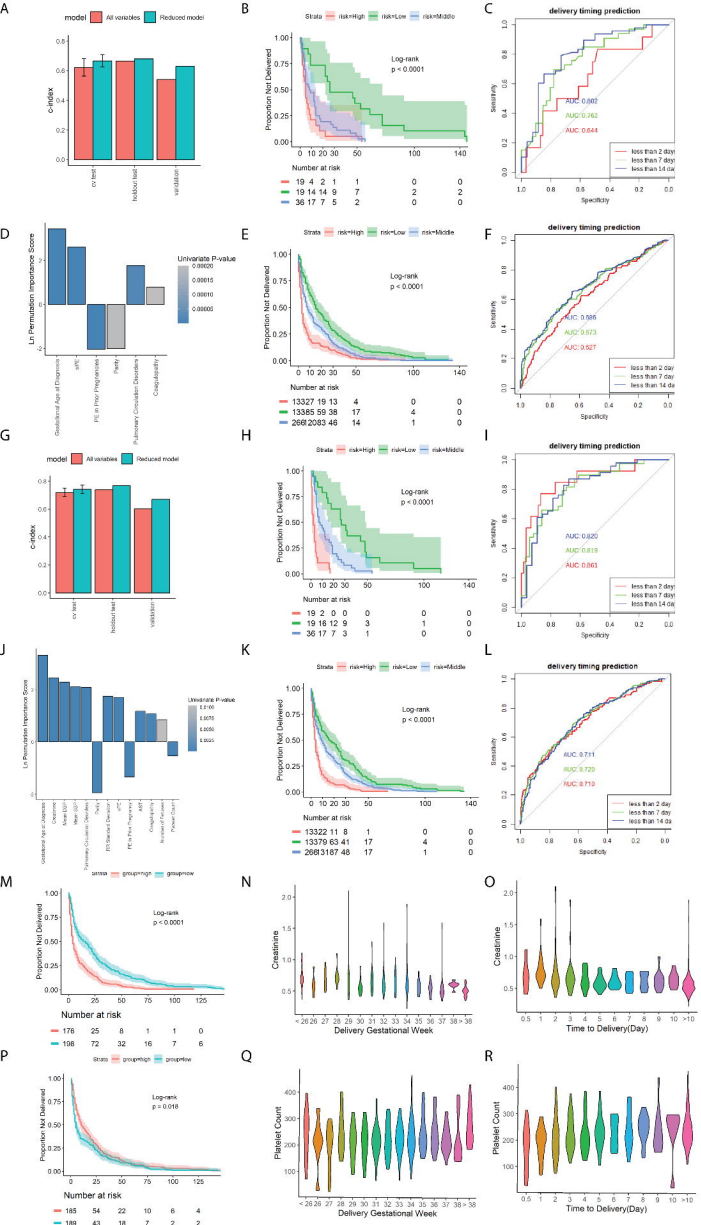
Survival Curves



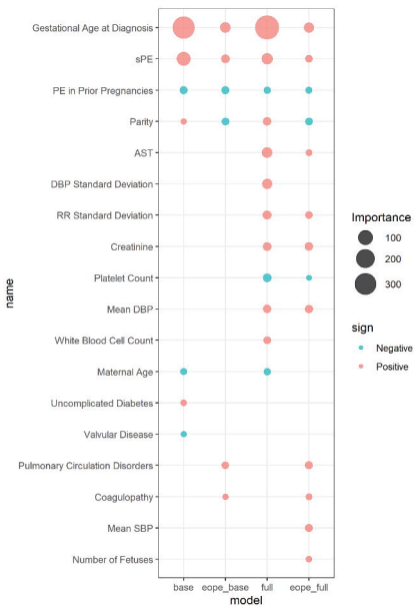
Shiny App







A



B

