

Prediction Model for Pre-Eclampsia Using Gestational Age Specific Serum Creatinine

Jieun Kang ¹, Sang Won Hwang ², Taesic Lee ³, Jooyoung Cho ⁴, Dong Min Seo ⁵, Seong Jin Choi ^{1,*}, Young Uh ^{4,*}

¹ Department of Obstetrics and Gynecology, Yonsei University Wonju College of Medicine, Wonju 26411, Republic of Korea

² Artificial Intelligence Bigdata Medical Center, Yonsei University Wonju College of Medicine, Wonju 26411, Republic of Korea

³ Department of Family Medicine, Yonsei University Wonju College of Medicine, Wonju 26411, Republic of Korea

⁴ Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju 26411, Republic of Korea

⁵ Department of Medical Information, Yonsei University Wonju College of Medicine, Wonju 26411, Republic of Korea

* Correspondence: choisj@yonsei.ac.kr; u931018@yonsei.ac.kr; Tel.: +82-33-741-1592

1. Definitions of adverse pregnancy outcomes

PE was defined as the presence of HTN and proteinuria occurring after 20 weeks of gestation in a previously normotensive patient, based on the American College of Obstetrics and Gynecology [17,18].

PTB was determined as birth before 37 completed weeks of gestation. This definition is recommended by the World Health Organization and is used in the International Classification of Diseases. Infants born toward the end of this preterm period were traditionally considered to have “lower risk” [19]. Therefore, babies born between 34+0 weeks and 36+6 weeks of GA were defined as late preterm infants. Moreover, PTB was subdivided into early PTB (~ 33+6 weeks) and late PTB (34+0 to 36+6 weeks).

FGR was defined as a failure to achieve the appropriate growth potential. Fetuses with birth weight below the 10th percentile are at an increased risk of stillbirth or perinatal mortality, while those with birth weight below the 3rd percentile have the highest risk. At present, there is no gold standard for the diagnosis of FGR. Therefore, it is usually defined as the statistical deviation of the fetal size from a population-based reference, typically at thresholds in the 10th, 5th, or 3rd centile and most commonly in the 10th centile [20,21].

| Serum Creatinine data | | | | | Anthropometric data | | | |
|-----------------------|------|------|--------|-----|---------------------|--------|-------|----------|
| id | date | mt | cr | GAQ | id | wt | ht | bmi |
| 1 | 2010 | 0.29 | 79.560 | 0 | 1 | 106.00 | 162.0 | 40.39018 |
| 1 | 2010 | 0.14 | 79.560 | 0 | 1 | 67.60 | 160.7 | 26.17670 |
| 1 | 2012 | 0.43 | 44.200 | 0 | 1 | 80.20 | 170.0 | 27.75087 |
| 1 | 2011 | 0.14 | 61.880 | 0 | 1 | 76.00 | 169.3 | 26.51549 |
| 1 | 2011 | 0.14 | 61.880 | 0 | 1 | 76.00 | 167.0 | 27.25089 |
| 1 | 2014 | 0.29 | 61.880 | 0 | 1 | 62.00 | 160.0 | 24.21875 |
| 1 | 2014 | 0.86 | 50.388 | 0 | 1 | 47.80 | 151.0 | 20.96399 |
| 1 | 2016 | 0.29 | 66.300 | 0 | 1 | 60.50 | 162.5 | 22.91124 |
| 1 | 2010 | 0.14 | 53.040 | 0 | 1 | 59.00 | 162.0 | 22.48133 |
| 9 | 2013 | 0.71 | 67.184 | 0 | 1 | 62.45 | 168.7 | 21.94332 |
| 1 | 2016 | 0.14 | 63.648 | 0 | 1 | 63.60 | 163.0 | 23.93767 |
| 1 | 2012 | 0.43 | 57.460 | 0 | 1 | 61.00 | 156.0 | 25.06575 |
| 1 | 2013 | 0.14 | 62.764 | 0 | 1 | 76.00 | 162.0 | 28.95900 |
| 9 | 2011 | 0.57 | 61.880 | 0 | 1 | 80.00 | 162.0 | 30.48316 |
| 1 | 2019 | 0.29 | 60.996 | 0 | 1 | 61.80 | 165.0 | 22.69972 |
| 1 | 2011 | 0.57 | 53.040 | 0 | 1 | 108.00 | 176.1 | 34.82612 |
| 1 | 2011 | 0.86 | 61.880 | 0 | 1 | 57.00 | 163.0 | 21.45357 |
| 8 | 2015 | 0.29 | 65.416 | 0 | 1 | 75.00 | 163.0 | 28.22839 |
| 1 | 2018 | 0.43 | 60.996 | 0 | 1 | 90.00 | 165.0 | 33.05785 |
| 1 | 2013 | 0.86 | 67.184 | 0 | 2 | 47.00 | 153.0 | 20.07775 |

| Lab data | | | | | | | |
|------------|-----|-----|-----|-----|------|-----|-----------|
| patient.no | alp | ast | alt | ldh | bun | ggt | test.date |
| 1 | 96 | 15 | 9 | | 4 | 16 | 201 |
| 1 | 147 | 77 | 59 | | 5 | 28 | 201 |
| 9 | | 57 | 31 | 379 | 26.8 | | 201 |
| 1 | 75 | 33 | 15 | | 14 | 7 | 201 |
| 1 | | 15 | 6 | | | 12 | 201 |
| 9 | 59 | 22 | 22 | | 14.2 | 15 | 202 |
| 2 | 114 | 58 | 30 | 292 | 8.9 | 71 | 202 |
| 1 | 73 | 21 | 8 | | 11.3 | 11 | 201 |
| 1 | 62 | 20 | 12 | | 17 | 5 | 201 |
| 1 | | 20 | < 6 | 283 | 12.5 | | 201 |
| 1 | | 18 | 11 | 221 | 8.1 | | 201 |
| 1 | 61 | 17 | 26 | 202 | 11.1 | 28 | 201 |

Figure S1. Data structure

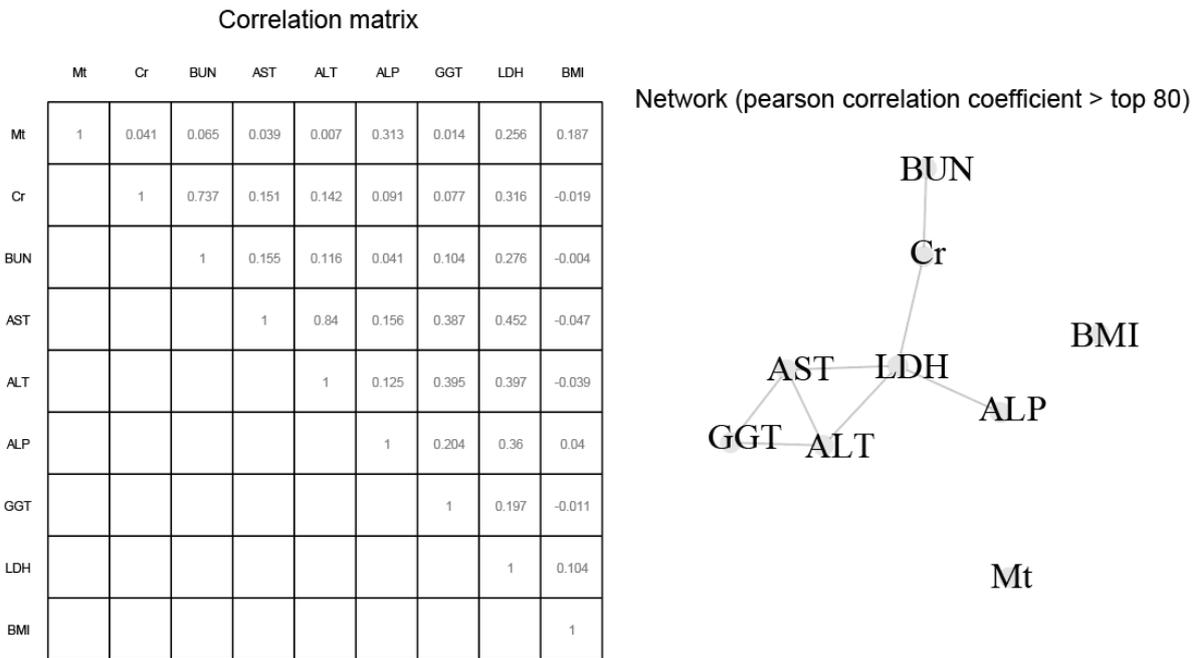


Figure S2. Correlation analysis among candidate markers.

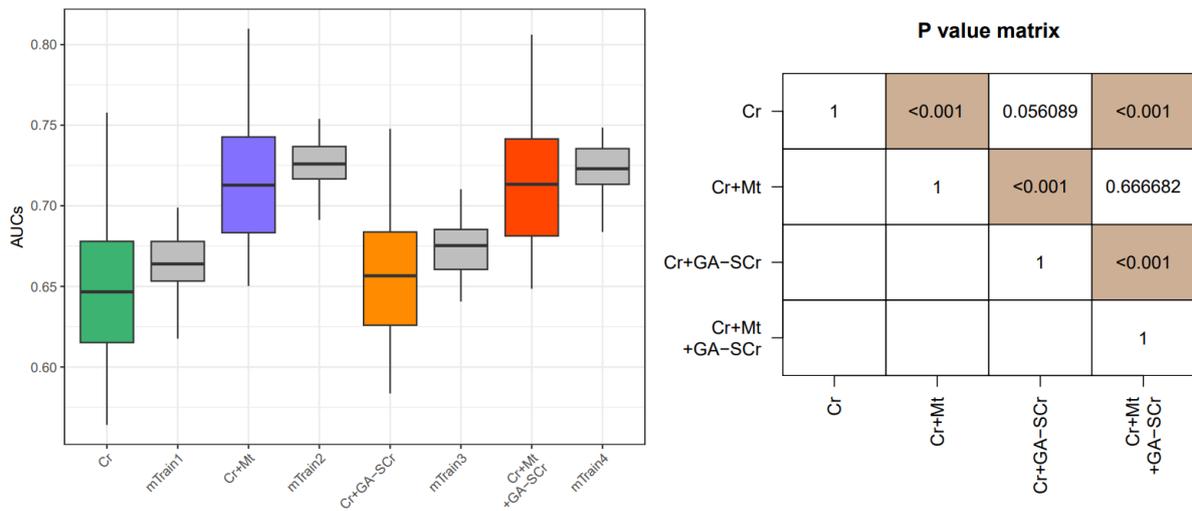
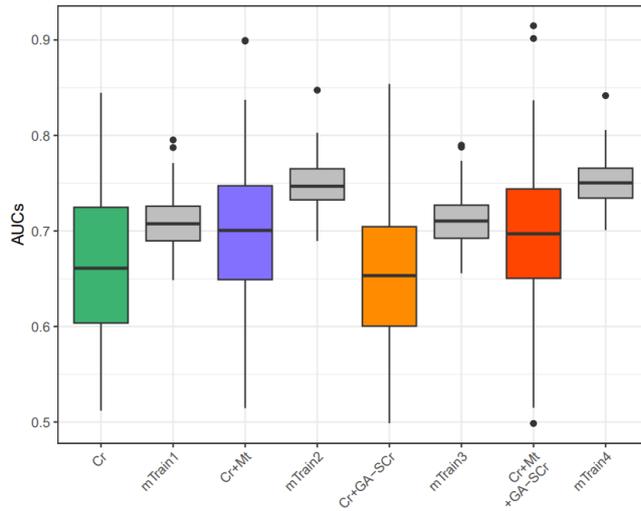


Figure S3. Prediction performances of PE with late PTB using the four feature-sets. Any cases of PE (PE only + PE with other adverse pregnancy outcomes) were determined as the disease group. Prediction model was established using logistic regression. Boxplots include minimum, first quartile (25%), median, third quartile, and maximum values. Points in the boxplot referred to as outliers indicate cases showing more than 1.5 times the interquartile range (IQR), biased from the matched median value. Y-axis indicates AUC for predicting PE, and the box plot summarizes 100 levels of AUCs. Green, purple, orange, and red colored boxplots were obtained from the testing dataset while grey colored boxplots are curated from the training dataset. Abbreviations: PE, preeclampsia; SCr, serum creatinine; Mt, measured time; GAQ, gestational age-specific SCr quartile; IQR, interquartile range; AUC, area under receiver characteristic curve.



P value matrix

| | | | | |
|--------------|----|----------|-----------|--------------|
| Cr | 1 | 0.000799 | 0.577104 | 0.002372 |
| Cr+Mt | | 1 | 8.3e-05 | 0.73675 |
| Cr+GA-SCr | | | 1 | 0.00029 |
| Cr+Mt+GA-SCr | | | | 1 |
| | Cr | Cr+Mt | Cr+GA-SCr | Cr+Mt+GA-SCr |

Figure S4. Prediction performances of PE with FGR using the four feature-sets. Any cases of PE (PE only + PE with other adverse pregnancy outcomes) were determined as the disease group. Prediction model was established using logistic regression. Boxplots include minimum, first quartile (25%), median, third quartile, and maximum values. Points in the boxplot referred to as outliers indicate cases showing more than 1.5 times the interquartile range (IQR), biased from the matched median value. Y-axis indicates AUC for predicting PE, and the box plot summarizes 100 levels of AUCs. Green, purple, orange, and red colored boxplots were obtained from the testing dataset while grey colored boxplots are curated from the training dataset. Abbreviations: PE, preeclampsia; SCr, serum creatinine; Mt, measured time; GAQ, gestational age-specific SCr quartile; IQR, interquartile range; AUC, area under receiver characteristic curve.

Table S1. General participant characteristics according to the GAQ.

| | GAQ1 | GAQ2 | GAQ3 | GAQ4 | P for trend |
|---------------------------|--------------|--------------|--------------|---------------|-------------|
| N | 5847 | 1794 | 1517 | 968 | |
| Age | 33.17 ± 0.07 | 33.65 ± 0.11 | 33.56 ± 0.13 | 33.6 ± 0.16 | <0.001 |
| Labor types, n | | | | | |
| Nullipara | 2595 (44.4) | 696 (38.8) | 552 (36.4) | 362 (37.4) | <0.001 |
| Multipara | 3252 (55.6) | 1098 (61.2) | 965 (63.6) | 606 (62.6) | <0.001 |
| Essential hypertension, n | 72 (1.2) | 17 (0.9) | 31 (2) | 33 (3.4) | <0.001 |
| Diabetes, n | 188 (3.2) | 75 (4.2) | 54 (3.6) | 90 (9.3) | <0.001 |
| PE, n | 461 (7.9) | 221 (12.3) | 243 (16) | 291 (30.1) | <0.001 |
| PE alone, n | 75 (1.3) | 37 (2.1) | 40 (2.6) | 17 (1.8) | 0.001 |
| PE+FGR, n | 37 (0.6) | 25 (1.4) | 19 (1.3) | 33 (3.4) | <0.001 |
| PE+PTB, n | 151 (2.6) | 79 (4.4) | 75 (4.9) | 127 (13.1) | <0.001 |
| PE+FGR+PTB, n | 198 (3.4) | 80 (4.5) | 109 (7.2) | 114 (11.8) | <0.001 |
| SCr, µmol/L | 40.35 ± 0.11 | 53.01 ± 0.13 | 60.11 ± 0.17 | 116.18 ± 3.21 | <0.001 |
| SCr < 30 | 575 | 0 | 0 | 0 | |
| 30 ≤ SCr <40 | 2109 | 0 | 0 | 0 | |
| 40 ≤ SCr <50 | 2537 | 567 | 2 | 0 | |
| 50 ≤ SCr <60 | 600 | 942 | 799 | 14 | |
| 60 ≤ SCr <110 | 26 | 285 | 716 | 718 | |
| eGFR ≥ 110 | 0 | 0 | 0 | 236 | |

Continuous and categorical variables are presented as mean ± SD and number (percent), respectively. Abbreviations: SCr, serum creatinine; SD, standard deviation; PE, preeclampsia; FGR, fetal growth retardation; PTB, preterm birth.