Peer Review File

Article information: https://dx.doi.org/10.21037/tp-22-611

Reviewer A

General comments

This study used a huge developing cohort in USA, and validated the models using relatively large cohort in the different areas. Sample size and study design is excellent. They identified many risk factors for preterm birth with small confidence intervals. In addition, they constructed nomograms to predict the individual probability of preterm birth, which will be useful in real medicine.

Reply: Thank you for your appreciation. All of your words and work are highly appreciated.

Comments

Introduction

The authors clearly stated the current understanding and unknown clinical questions on prediction models for preterm birth.

Reply: Thank you for your appreciation. All of your words and work are highly appreciated.

Materials and Methods

You listed relevant variables. What do you mean “hypertension eclampsia”? Because you listed hypertension (pre-pregnancy hypertension, gestational hypertension, and no hypertension) in the lists, does the “hypertension eclampsia” indicate both pre-eclampsia and eclampsia? OR, does the “hypertension eclampsia” indicate only eclampsia without showing hypertension? If so, where is the category of preeclampsia or super-imposed preeclampsia included in?

Reply: We are very sorry that our expression confused you. The hypertensive eclampsia in this paper is a convulsion that occurs in the mother on the basis of preeclampsia and cannot be explained by other

reasons. Therefore, in this paper, the hypertension eclampsia only indicate eclampsia. Preeclampsia is also associated with preterm birth in previous study. Unfortunately, however, this database did not collected the data on preeclampsia. So we did not included this factor of preeclampsia in our models, and that is one of the limitations of our study.(see Page 9, line 240)

Line 103: full-term birth (FTB) -> FTB

Reply: Thank you for your reminding.

Changes in the text: We have modified the full-term birth (FTB) to FTB in the paper. (see Page 3, line 86)

There are no concerns on the statistical methods. However, you should add how to use nomogram.

Reply: Thanks for your comments for our statistical methods. We add the nomogram usage method in this paper. (see Page 5, line 130-134)

Changes in the text: To use the nomogram, first, locate a specific point of an individual patient on each variable axis; second, draw a vertical line upwards in variable axis, each variable corresponding to a specific point on the Points scale (top); third, sum all the points of each variable, and locate on the Total Points scale (bottom); forth, draw a vertical line downwards, corresponding to the PTB probability axis to determine the PTB probability. (see Page 5, line 130-134)

Results

Table 1, Supplementary Table S1: You should list up “FTB” in the foot note, presenting the full spelling.

Reply: Thanks for your suggestions. we added the full spelling of FTB in the foot note of Supplementary Table S1. (see Page 1 in Supplementary materials)

Changes in the text: The full spelling of FTB was added in the foot note of Supplementary Table S1. (see Page 1 in Supplementary materials)

You should also make the Supplementary Tables showing the score of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points, respectively.

Reply: Thanks for your suggestions. we added the score of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points, respectively, in Supplementary Table S6, S7 and S8. (see Page 11-13 in Supplementary materials)

Changes in the text: The score of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points, respectively, were added in Supplementary Table S6, S7 and S8. (see Page 11-13 in Supplementary materials)

Discussion

Line 223 thees -> these or the?

Reply: Thank you for your reminding. We are sorry that this is our clerical error.

Changes in the text: We have corrected it in the article. (see Page 7, line 187)

This study used a huge developing cohort in USA, and validated the models using

relatively large cohort in the different areas. Sample size and study design is excellent. They identified many risk factors for preterm birth with small confidence intervals. In addition, they constructed nomograms to predict the individual probability of preterm birth, which will be useful in real medicine.

However, several modification, especially individual scores for of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points should be added as supplementary tables.

Reply: Thanks for your suggestions. we added the score of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points, respectively, in Supplementary Table S6, S7 and S8. (see Page 11-13 in Supplementary materials)

Changes in the text: The score of each variable in the moderately and late PB, VPB, and EPB risk model and the prediction probability corresponding to the total points, respectively, were added in Supplementary Table S6, S7 and S8. (see Page 11-13 in Supplementary materials).

Reviewer B

Figure 3, we failed to find figures 3M-3P in figure 3, while they are mentioned in the figure 3 legend, please check it.

Figure 3. Time-dependent ROC curves of preterm birth (A), moderately and late preterm birth (B), very preterm birth (C), and extremely preterm birth (D) in training data; and time-dependent ROC curves of preterm birth (E), moderately and late preterm birth (F), very preterm birth (G), and extremely preterm birth (H) in independent validation data. Calibration curves for the risk of preterm birth (I), moderately and late preterm birth (J), very preterm birth (K), and extremely preterm birth (L) in training data; and calibration curves for the risk of preterm birth (M), moderately and late preterm birth (N), very preterm birth (O), and extremely preterm birth (P) in independent validation data. ROC, receiver operating characteristic; AUC, area under the ROC curve.

Response: We are very sorry that our clerical error confused you. We have made corresponding modifications.

Table 1 and main text, please change number format in thousands, for example, it should be presented as 14,518,114 rather than 14518114.

Response: Thanks for your suggestion, we have revised the relevant content.

Table 1, the percentages may not add to 100%, if relevant, authors may consider adding a footnote the explain why numbers do not sum to 100%. Please also check the whole text and ensure the accuracy of all numbers and values.

All population	14518114 (100)	1776415 (100)
Year		
2016	3021461 (20.8)	356284 (20.1)
2017	2954513 (20.4)	356739 (20.1)
2018	2913994 (20.1)	354401 (20.0)
2019	2873800 (19.8)	366091 (20.6)
2020	2754346 (19.0)	342900 (19.3)
Maternal age		
<20	528824 (3.6)	76366 (4,3)
20-30	6835689 (47.1)	785330 (44.2)
31-40	6679838 (46.0)	824490 (46.4)
>40	473763 (3.3)	90229 (5.1)

Response: We are very sorry that our clerical error confused you. As a result of the rounding method used to calculate the percentages, the percentages add up to more than 100.0%. We have carefully reviewed all numbers and values in our study, and the previous tables has been replaced by the new tables (Table1 and TableS1). In this version, we confirmed the whole text and all numbers and values were correct.

Table 2, please check the number circled to see if any adjustments should be made, as it may not be consistent with the main text.

Maternal age (year)			
< 20 vs 20-30	1.155	1.144-1.166	<0.001

Table1. Univariate analysis ~~found~~ showed that 20 variables had relationship with PTB. Table 2 lists ~~listed~~ the multivariate logistic regression analysis results, ~~showed~~ showing that the male newborn (OR: 0.910; 95% CI: 0.907-0.913; *P*<0.001), maternal age>40 (OR: 1.155; 95% CI:1.144-1.166; *P*<0.001), maternal BMI (OR: 1.224; 95% CI:1.212-1.236; *P*<0.001), marital status (OR: 1.142 ; 95% CI:1.137-1.146; *P*<0.001), mother's race (OR: 1.425; 95% CI: 1.418-1.432; *P*<0.001), mother's

Response: We are very sorry that our clerical error confused you. We have made corresponding modifications. The modified content is as follows: maternal age~~>40~~ <20 (OR: 1.155; 95% CI:1.144-1.166; *P*<0.001),

Any abbreviations used in figures and tables or their description should be defined in a footnote beneath each corresponding table/figure. Even if they were explained in the main text, full terms must be presented again in the corresponding figures and tables, so that figures and tables can be read on their own.

Response: Thanks for your suggestion, we have added footnotes under each figure and each table.

The mentioned name does not match with the authors name, please check it.

A similar PTB nomogram was ~~constructed~~ developed by Lee et al. with 192,208 samples[6], and they established their nomogram based on 9 factors. Our results supported their conclusion that these

6. Mehta-Lee SS, Palma A, Bernstein PS, Lounsbury D, Schlecht NF. A Preconception Nomogram to Predict Preterm Delivery. *Matern Child Health J* 2017; **21**(1): 118-127.

Response: We are very sorry that our clerical error confused you. The reference quoted here is correct. We have added the author's full name in the study.

Authors mention 'studies', while only one reference is cited, please check if add adjustments should be made.

193 between chorioamnionitis and gonorrheae[25]. Chorioamnitis was an independent predictor in previous

194 studies[26], and chlamydia and gonorrhea might lead to chorioamnionitis to increase PTB risk. And Gao

Response: Thank you very much for your reminding and suggestion, we have revised it in the manuscript.