Association of preterm birth with maternal heart disease : Machine learning analysis using national health insurance data

Eun Saem Choi[1*], Jue Seong Lee [2*], Yujin Hwang [1,3], Ho Yeon Kim[1], Geum Joon Cho[1], Soon Cheol Hong[1], Min Jeong Oh[1], Hai Joong Kim[1], Kwang-Sig Lee[3,] Ki Hoon Ahn[1**]

1 Department of Obstetrics and Gynecology, Korea University College of Medicine, Korea University Anam Hospital, Seoul, Korea
2 Department of Pediatrics, Korea University College of Medicine, Korea University Anam Hospital, Seoul, Korea
3 AI Center, Korea University College of Medicine, Korea University Anam Hospital, Seoul, Korea

Objective

The number of pregnant women with heart disease (HD) is rising along with the increased prevalence of underlying diseases such as hypertension, diabetes and obesity. In addition, increasing number of women with congenital heart disease (CHD) are surviving and able to reach reproductive age. It is noteworthy that the risk factors for heart diseases (HD) including age, socioeconomic status, or diabetes are also associated with preterm birth (PTB). However, previous studies were mostly focused on the cardiovascular risk of pregnant women with HD, not the risk of PTB. *The aim of this study was to assess the association between maternal HD and preterm birth by using machine learning and population data*.

Method

This population-based retrospective cohort study included 174,926 primiparous women who aged 25 - 40 and delivered in 2017. The 36 independent variables regarding demographic/socioeconomic factors and medical history, including known contributing factors of PTB were extracted from Korea National Health Insurance claims with ICD-10 and ATC codes. Maternal HD included as follows; CHD, arrhythmia, cardiomyopathy, congestive heart failure (CHF), ischemic heart disease (IHD), endocarditis, cardiac arrest. Machine learning analysis was used to establish the prediction model of PTB and random forest variable importance was used to identify the association between various maternal HD and PTB. To improve the explanation on how each value works in the prediction model, SHAP (Shapley additive explanations) value was computed.

Results

Among the study population, 12,701 (7.26%) had preterm birth and 12,234 women had at least one of the pregestational heart diseases. *The areas under the operating-characteristic-curve of prediction model using the random forest with oversampling data was 0.885 and the accuracy was 0.896.*

Based on the prediction model, maternal HD is associated with PTB. *Based on the feature importance of prediction model, PTB has strong associations with arrhythmia (feature importance, 15th, 0.0152) and IHD (feature importance, 17th, 0.0107) among maternal heart diseases (Table 1).* The variable importance of arrhythmia and IHD were higher than that of hypertensive disorders during pregnancy which is a well-known risk factor of PTB. The association between CHD (feature importance, 26th, 0.0008) and PTB were found to be lower than those of arrhythmia and IHD.

Based on SHAP value, IHD is associated with the increased the risk of PTB more consistently than arrhythmia (Fig. 1). To further explain the effect of arrhythmia on PTB, subgroup analysis of arrhythmia was performed. According to the subgroup analysis, atrial fibrillation/flutter was associated with the increased risk of PTB, while supraventricular tachycardia was not (Fig. 2).

Table 1. Random forest feature importance of prediction model for PTB (Top 20 variables)

Rank	Variables	Feature importance
1	Socioeconomic status	0.3377
2	Age	0.2881
3	Gestational diabetes	0.0391
4	Anemia	0.0329
5	Sepsis	0.0311
6	Abnormal menstruation	0.0285
7	Benzodiazepine	0.0249
8	Tricyclic antidepressant	0.0221
9	Progesterone	0.0214
10	Pregestational hypertension	0.0213
11	Vaginitis	0.0211
12	Hyperlipidemia	0.0186
13	Pelvic inflammatory disease	0.0184
14	Recurrent miscarriage or infertility	0.0162
15	Arrhythmia	0.0152
16	Hypnotic/sedative drug	0.0124
17	Ischemic heart disease	0.0107
18	Endometriosis	0.0094
19	Diabetes	0.0090
20	Hypertensive disorders during pregnancy	0.0084

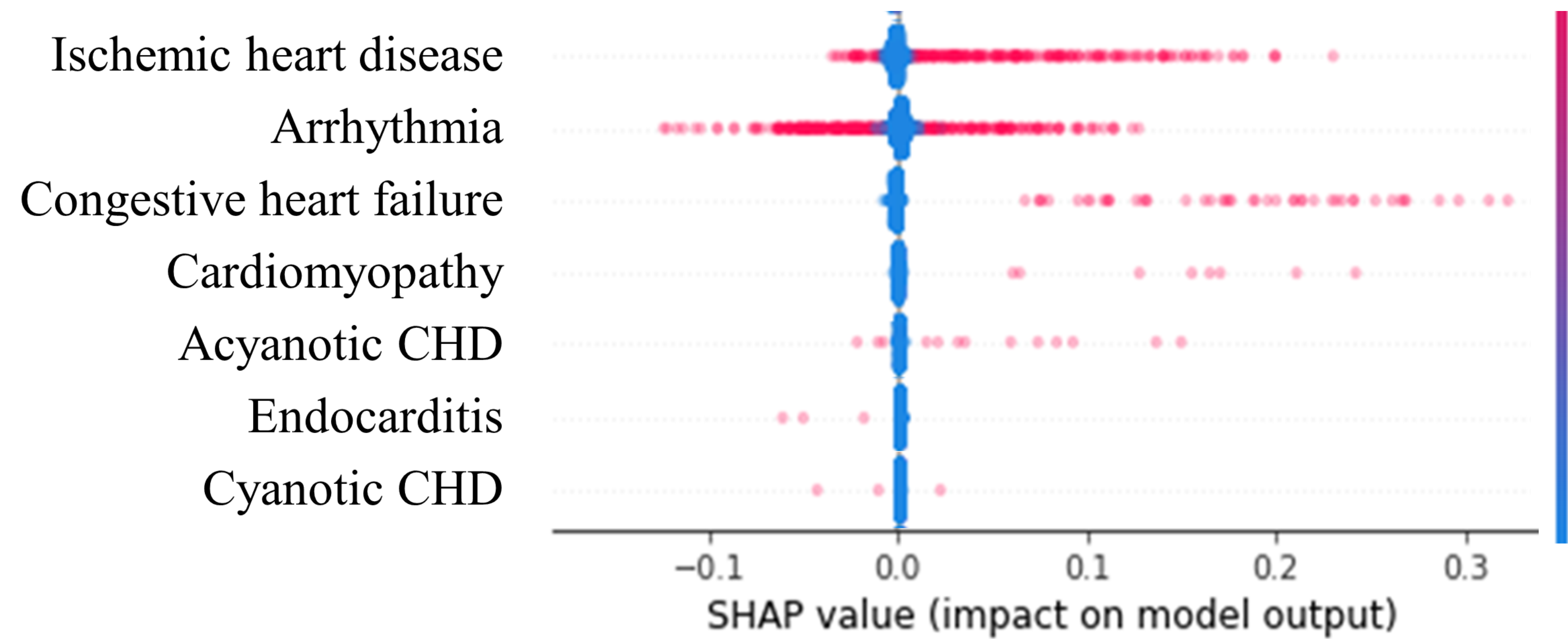**Figure 1. SHAP value of maternal heart diseases in prediction model for PTB**

Figure 2. SHAP value of arrhythmia subgroup in prediction model for PTB

Conclusion

By using machine learning, we established a valid prediction model for PTB and found that maternal HD, especially arrhythmia and IHD were associated with the increased risk of PTB. In addition, by SHAP value, we found that among various arrhythmia, atrial fibrillation/atrial flutter was the main cause of increased risk of PTB. This study implies that close surveillance integrating both cardiovascular and obstetric risk is needed in pregnant women with HD.

고려대학교 의과대학
Korea University College of Medicine