

Machine-Learning Prediction Model for Gestational Diabetes in Twin Pregnancies: Population-Based Cohort Study

Hiba J. Mustafa, Erkan Kalafat, Mohammad-Hossein Heydari, Rebecca N. Nunge, Asma Khalil.

Indiana University School of Medicine, USA; Koc University School of Medicine, Turkey; University of Medical Sciences, Iran; St George's University of London, UK

Objective

To assess prevalence of and risk factors for gestational diabetes (GDM) in twin compared with singleton pregnancies, and to develop a prediction model to quantify the risk of developing GDM in twin pregnancies.

Methods: Population-based study using CDC birth data from 2016-2021. Higher-order pregnancies and pre-pregnancy diabetes were excluded. A Chi-square test of independence was performed to identify factors significantly associated with GDM in twin versus singleton pregnancies, and within each group independently. Multivariable regression analyses were performed, first to assess the association between GDM and candidate predictors, and second to assess risk of GDM in twin compared with singletons, adjusted for the significant risk factors. All potential predictors were evaluated in two prediction models: logistic regression and machine-learning (ML). The predictive accuracy was assessed by AUROC. Non-parametric test was used to compare AUROC of both models and, given the large sample size, P value<0.001 was considered statistically significant.

Results

A total of 18,173,365 singleton and 611,043 twin pregnancies were included. Following the regression model, maternal age $\geq$ 30 years, nulliparity, IVF, chronic hypertension, Hispanic and Non-Hispanic (NH) Asian, foreign-born, overweight and obesity class I/II/III remained significantly associated with GDM in twins. However, maternal age<25 years and W.I.C program reduced that risk. Factors that more than doubled the risk in twins were maternal age $\geq$ 40 years (OR 2.06 (1.97 – 2.14), $P<0.001$), NH Asian (OR 2.12 (2.04 – 2.20), $P<0.001$), and obesity class I, II, and III (OR: 2.22 (2.16 – 2.29), $P<0.001$; OR: 3.01 (2.92 – 3.11), $P<0.001$; OR: 3.80 (3.67 – 3.93), $P<0.001$, respectively). Following adjustment for all significant risk factors, twin pregnancy remained significantly associated with increased risk of GDM in twin compared to singleton pregnancies (OR 1.22 (1.21 – 1.23), $P<0.001$). Predictors incorporated in the two prediction models were age, parity, race-ethnicity, nativity, education, BMI, insurance, Women, Infants and Children (W.I.C) program, peri-conception cigarette use, chronic hypertension, and IVF. The prediction performance of the ML model was similar to that of logistic regression, with AUROC of 0.681 (95% CI: 0.678 to 0.684) vs. 0.679 (95% CI: 0.676 to 0.682) ($P=0.257$). Both models were well-calibrated and none had poor fit. An online calculator was developed for risk prediction.

Conclusions

Of the significant risk factors, maternal age $\geq$ 40 years, race-ethnicity, and obesity more than doubled the risk of GDM in twin compared to singleton pregnancy. The prediction model of GDM, over and above that in the general twin population.