Predicting cesarean delivery with decision tree models

Cynthia J. Sims, MD,[a] Leslie Meyn, BS,[a] Rich Caruana, PhD,[b, c] R. Bharat Rao, PhD,[d] Tom Mitchell, PhD,[b] and Marijane Krohn, PhD[a]

Pittsburgh, Pennsylvania, Los Angeles, California, and Princeton, New Jersey

OBJECTIVE: The purpose of this study was to determine whether decision tree–based methods can be used to predict cesarean delivery.

STUDY DESIGN: This was a historical cohort study of women delivered of live-born singleton neonates in 1995 through 1997 (22,157). The frequency of cesarean delivery was 17%; 78 variables were used for analysis. Decision tree rule-based methods and logistic regression models were each applied to the same 50% of the sample to develop the predictive training models and these models were tested on the remaining 50%.

RESULTS: Decision tree receiver operating characteristic curve areas were as follows: nulliparous, 0.82; parous, 0.93. Logistic receiver operating characteristic curve areas were as follows: nulliparous, 0.86; parous, 0.93. Decision tree methods and logistic regression methods used similar predictive variables; however, logistic methods required more variables and yielded less intelligible models. Among the 6 decision tree building methods tested, the strict minimum message length criterion yielded decision trees that were small yet accurate. Risk factor variables were identified in 676 nulliparous cesarean deliveries (69%) and 419 parous cesarean deliveries (47.6%).

CONCLUSION: Decision tree models can be used to predict cesarean delivery. Models built with strict minimum message length decision trees have the following attributes: Their performance is comparable to that of logistic regression; they are small enough to be intelligible to physicians; they reveal causal dependencies among variables not detected by logistic regression; they can handle missing values more easily than can logistic methods; they predict cesarean deliveries that lack a categorized risk factor variable. (Am J Obstet Gynecol 2000;183:1198-206.)

Key words: Decision trees, machine learning, predicting cesarean delivery, statistical models