Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models

Ali Al Housseini, MD; Tondra Newman, MD; Alan Cox, MBA; Lawrence D. Devoe, MD

OBJECTIVE: We sought to develop a neural network (NN) to predict the risk for cesarean delivery (CD) in term nulliparas.

STUDY DESIGN: Using software (BrainMaker for Windows, Version 3.0; California Scientific Software, Nevada City, CA), we trained an NN with 225 patients obtained by chart review and included for nulliparity, singleton vertex > 36 weeks' gestation, and reassuring fetal heart rate on admission. Training inputs included several maternal and fetal clinical variables. Two logistic regression (LR) models using 225 and 600 patients (LR225 and LR600, respectively) were developed. The NN and LR models were tested for prediction of CD in a set of 100 patients not used for development.

RESULTS: The NN, LR225, and LR600 correctly predicted 53%, 26%, and 32% of the patients with CD and 88%, 95%, and 95% of the patients with vaginal delivery, respectively.

CONCLUSION: Compared with LRs, the NN was slightly better in predicting CD and was similar for predicting vaginal delivery in nulliparas with term singletons.

Key words: cesarean, neural network, prediction, vaginal

Cite this article as: Al Housseini A, Newman T, Cox A, et al. Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models. Am J Obstet Gynecol 2009;201:113.e1-6.

Cesarean delivery (CD) has increased by > 350% in the United States since the 1970s, reaching a rate of 29.1% in 2004.[1,2] An American College of Obstetricians and Gynecologists (ACOG) Task Force, constituted in 1999 to study the growing CD "epidemic," found that one of the greatest drivers of higher CD rates came from the group of nulliparas with singleton term gestations.[3] Although their report made recommendations for limiting CDs in this group, these recommendations appear to have had little, if any, impact on subsequent CD rates.

From the Medical College of Georgia, Augusta, GA.

Presented as a topic of the John McCain Presentation at the 71st Annual Meeting of the South Atlantic Association of Obstetricians and Gynecologists, Asheville, NC, Jan. 17-20, 2009.

Received Oct. 1, 2008; revised March 11, 2009; accepted May 1, 2009.

Reprints not available from the authors.

This study was supported by the Department of Obstetrics and Gynecology, Medical College of Georgia, Augusta, GA.

0002-9378/$36.00

doi: 10.1016/j.ajog.2009.05.001

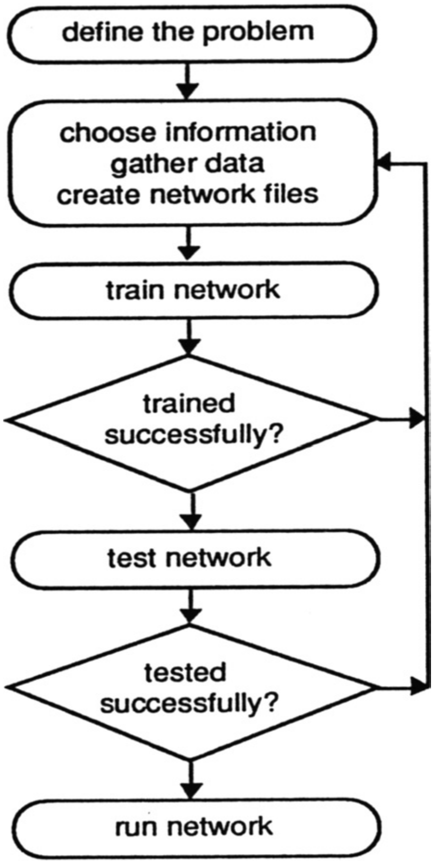
Among the concerns associated with labor in term nulliparas is the unpredictability of its course and outcome. This concern may also contribute to another practice phenomenon, CD on maternal request. A National Institute of Child Health and Human Development (NICHD) expert panel studied this issue and concluded that, in the absence of any increase in known clinical risk factors for primary CD, CD on maternal request may be contributing to higher rates of primary CDs.[2] In view of such concerns, it would be reasonable to raise the following question: would CD be performed more appropriately and would some of these maternal concerns be mitigated if labor outcome could be predicted more accurately in advance?

Previous attempts to develop models that predict the mode of delivery, based on traditional formulaic and statistical approaches, have incorporated variables such as Bishop scores and other maternal or fetal variables.[4-10] To date, these models have shown limited predictive accuracy. Artificial neural networks (NNs) are "intelligent" software programs that have been used successfully to solve obstetric and gynecologic prediction problems such as breast cancer survival, estimation of fetal weight, duration of active labor, and the significance of fetal heart rate (FHR) patterns.[11-18] Given the importance of predicting the outcome of labor in a group of patients with no obstetric history, better prediction models could prove beneficial to obstetric providers and their patients. Figure 1 shows the building of the NN from scratch starting with defining a problem or a question, choosing the information, and gathering data and creating network files that will train the network. If the network gets trained successfully, it gets tested with a sample with a known outcome. If the network gets tested successfully, it is run to predict the outcome of new patients. If the network does not get trained or tested successfully the question or the input characteristics are modified and the network is tested again. Therefore, our goal was to develop an NN to predict the mode of delivery in term nulliparas more accurately than might be accomplished with conventional modeling, such as logistic regression (LR).

Materials and Methods

We reviewed the completed medical records of all obstetric patients who delivered at our institution between 2005 and 2007. This review was approved by

FIGURE 1
Neural network



Neural network development.

Al Housseini. Prediction of risk for CD in term nulliparas. Am J Obstet Gynecol 2009.

our institutional review board prior to its inception. We identified 700 term nulliparas who met the following inclusion criteria: (1) singleton pregnancy; (2) vertex presentation; (3) adequate pregnancy dating; (4) reassuring FHR pattern on admission; and (5) gestational age > 36 weeks. Patients were excluded for the following reasons: (1) active herpes infection; (2) placenta previa; (3) placental abruption; (4) suspected fetal growth restriction; (5) severe preeclampsia; (6) chorioamnionitis; and (7) rupture of membranes > 18 hours. Of these 700 patients, 600 patients were selected to create the NN and LR models and 100 patients were randomly segregated by the NN program to form a set for testing the resulting prediction models.

The NN was developed with software (BrainMaker for Windows, version 3.0; California Scientific Software, Nevada City, CA). The NN structure consisted of 3 layers, feed-forward with a back propagation algorithm. The NN predicts the outcome based on input characteristics and improves its predictive ability and experience by correlating the outcome with its input characteristics. The NN contained 1 input layer with 15 inputs, 1 hidden layer, and 1 output layer with 2 outputs (CD and vaginal delivery). The NN output is a probability between 0 and 1, similar in overall range to that of LR models. In our model, a probability that exceeded .5 would trigger a "yes" response for CD; a value below that threshold would trigger a "yes" response for vaginal delivery. This value is fixed at the start of NN development. The NN was developed and trained using a learning rate of 1.0 and tolerance of 0.1. The NN inputs included both maternal and fetal clinical variables. The maternal variables were age, weeks of gestation, height, weight, body mass index (BMI), systolic and diastolic blood pressure on admission, Bishop score on admission, membrane status on admission, group B beta-streptococcus carriage, ethnicity, and labor induction agents. Fetal variables were estimated fetal weight, station on admission, and baseline FHR. After a number of iterations, we determined that the NN was optimally trained with a set of 225 patients.

LR is a mathematical modeling approach that can be used to describe the relationship of several predictor variables ($X_1, X_2, \ldots, X_k$) to a dichotomous dependent variable Y. For our study, the predictor variables were age, agent for labor induction, weeks of gestation, estimated fetal weight, station, Bishop, FHR, height, weight, BMI, systolic and diastolic blood pressure, ethnicity, membrane status, and group B beta-streptococcus carriage. The dichotomous dependent variable was delivery mode. Continuous variables (eg, weeks of gestation, estimated fetal weight, station) were treated continuous while discrete variables (eg, induction, ethnicity, membrane status) were treated categorically. We used the Hosmer-Lemeshow goodness-of-fit test to assess linearity in assessing the log odds scale. We also tested for significant interactions among variables.

We developed 2 multiple regression models, LR225 and LR600, using the 225 patients in the NN training set and all 600 patients held out for NN development, respectively. The prediction performances of the NN and LR models were tested with a set of 100 patients who had not been used in model development. The distributions of the variables among the training and testing sets were examined using χ^2 test or 1-way analysis of variance, where appropriate, with an alpha error of 0.05. Ranking of the ability to predict delivery mode among the 3 different models was done using kappa statistics for agreement between predicted and actual outcomes.

Results

Patient characteristics for the NN and LR training sets and the testing set are shown in Table 1. We compared the effects of the additional 375 patients in the second LR model using analysis of variance. Increasing the size of the LR dataset resulted in significant increases in CDs compared with the original LR dataset and the testing set.

Table 2 shows the assessment of the individual variables used in LR model development. The only significant variables were Bishop score, BMI, weight, and age; only the interaction between these variables was calculated. Of the individual variables used for LR development, Bishop scoring was the best single predictor for both LR models. Using LR forward selection procedure ($P < .05$), the Bishop score was significantly selected for the 225 patient training set ($P = .0003$) and the 600 patient training set ($P = .0001$). Our results showed that only age × BMI was significant and this interaction term was used in the model.

The receiver operating characteristic curves for LR225 and LR600 are shown in Figures 2 and 3, respectively. The area under the receiver operating characteristic curve is used as a measure of diagnostic accuracy and can vary from 0.0-1.0. The areas under the receiver operating characteristic curve were