Two new mathematical models for prediction of early mortality risk in coronary artery bypass graft surgery

Alireza Alizadeh Ghavidel, MD,[a] Hoda Javadikasgari, MD,[a] Majid Maleki, MD,[b] Arsha Karbassi, MD,[c] Gholamreza Omrani, MD,[a] and Feridoun Noohi, MD[b]

Objectives: The aim of this study was to develop new models for prediction of short-term mortality risk in on-pump coronary artery bypass grafting (CABG) surgery using decision tree (DT) methods.

Methods: Between September 2005 and April 2006, 948 consecutive patients underwent CABG surgery at Rajaie Heart Center. Potential risk factors were reviewed and univariate and multivariate analysis for short-term mortality were performed. The whole dataset was divided into mutually exclusive subsets. An entropy error fuzzy decision tree (EEFDT) and an entropy error crisp decision tree (EECDT) were implemented using 650 (68.6%) patient data and tested with 298 (31.4%) patient data. Ten times hold-out cross validation was done and the area under the receiver operative characteristic curve (AUC) was reported as model performance. The results were compared with the logistic regression (LR) model and EuroSCORE.

Results: The overall short-term mortality rate was 3.8%, and was statistically higher in women than men ($P < .001$). The final EEFDT selected 19 variables and resulted in a tree with 39 nodes, 20 conditional rules, and AUC of 0.90 ± 0.008. The final EECDT selected 15 variables and resulted in a tree with 35 nodes, 18 conditional rules, and AUC of 0.86 ± 0.008. The LR model selected 10 variables and resulted in an AUC of 0.78 ± 0.008; the AUC for EuroSCORE was 0.77 ± 0.003. There were no differences in the discriminatory power of EEFDT and EECDT ($P = .066$) and their performance was superior to LR and EuroSCORE.

Conclusions: EEFDT, EECDT, LR, and EuroSCORE had clinical acceptance but the performance and accuracy of the DTs were superior to the other models. (J Thorac Cardiovasc Surg 2014;148:1291-8)

ACD

Supplemental material is available online.

Mortality risk evaluation has been increasingly emphasized in cardiac surgery.[1] The aims of developing risk models include quality monitoring of surgical performance, counseling patients and deciding treatment, cost-benefit analysis, or all of these purposes. Coronary artery bypass grafting (CABG) is one of the most common cardiac surgeries and most of the candidate patients are old with comorbidities. Thus, accurate prediction of operative risk is critical for patients and doctors to make a proper informed decision about surgery.[2]

From the Heart Valve Disease Research Center,[a] Rajaei Cardiovascular Medical and Research Center; Department of Cardiology,[b] Shahid Rajaie Heart Hospital; Interventional Cardiology,[c] Tehran Heart Center, Tehran, Iran.

Disclosures: Authors have nothing to disclose with regard to commercial support.

Received for publication Aug 4, 2012; revisions received Sept 1, 2013; accepted for publication Feb 3, 2014; available ahead of print March 7, 2014.

Address for reprints: Hoda Javadikasgari, MD, Heart Valve Disease Research Center, Rajaei Cardiovascular Medical and Research Center, Vali-Asr Ave, Niyayesh Blvd, Tehran 199691-1151, Iran (E-mail: hoda.javadi.k@gmail.com).

0022-5223/$36.00

http://dx.doi.org/10.1016/j.jtcvs.2014.02.028

Several cardiac surgery risk models have been proposed.[3,4] The EuroSCORE model has been shown to have the highest discriminatory power among all models; its area under the receiver operating characteristic (ROC) curve (AUC) did not exceed 0.78.[5]

Most risk scoring systems, like EuroSCORE, have been developed based on assumption of a linear relationship among variables. Therefore, there is a need to evaluate newer methods with more complex mathematical assumptions. Decision trees (DTs) are popular reasoning methods and have been successfully applied in clinical decision making. For example, a DT has recently been used to determine the most appropriate method of rectal cancer management.[6] Moreover, it has been shown that DTs improved the prediction of response to neoadjuvant therapy in breast cancer.[7] The main advantage of DTs is their interpretability. Moreover, their extracted conditional rules can be used for developing expert systems.

The aim of this study was to develop and compare the performance of DT models for prediction of short-term mortality risk in patients who undergo on-pump CABG surgery with or without concomitant valve replacement surgery. We also compared their results with EuroSCORE,[4] which is the most practical existing method in cardiac surgery risk assessment.

Abbreviations and Acronyms

ANN = artificial neural networks
AUC = area under the curve
CABG = coronary artery bypass grafting
DT = decision tree
EECDT = entropy error crisp decision tree
EEFDT = entropy error fuzzy decision tree
ICU = intensive care unit
LR = logistic regression
MAE = mean absolute error
ROC = receiver operating characteristic

PATIENTS AND METHODS

The protocol of this study was approved by the local ethical committee of the medical faculty at Tehran University of Medical Sciences. The need for informed consent was waived.

Patient Population and Data Collection

The dataset includes the information for 1068 consecutive patients who were referred for CABG surgery at Rajaie Cardiovascular Medical and Research Center between September 2005 and April 2006. The study was mainly designed to develop a regression model to predict mortality and assess the quality of medical care among different surgeons and institutions.[8] Patients with incomplete information on any predictive variable or outcome were excluded from this study. Postoperative variables were excluded from the multivariate analysis because new models can predict the preoperative mortality risk.

Outcome

The primary outcome was 30-day postoperative mortality. This included all patients who died either at the hospital or within 30 days of the operation date. Follow-up took place at an outpatient clinic or by telephone interview. Postoperative data were length of intensive care unit (ICU) stay and major complications after surgery including myocardial infarction, cardiac output state, prolonged ventilation, central nervous system complications, serious infections, and oliguria or anuria.

Statistical Analysis

Continuous variables are summarized by means $\pm$ standard deviation and categorical variables are expressed as proportions (%). Univariate analyses were performed by either χ^2 or Student t test where applicable. Model development and statistical analyses were performed using MATLAB software (V7.8, R2009A) and SPSS statistical software (version 15.0.0; SPSS Inc, Chicago, Ill).

Decision Tree Model

DT modeling is a popular reasoning method and has been successfully used in medical decision making. DTs consist of multilayer connected nodes and each branch of a DT from root node to a terminal node results in an individual conditional rule (Appendix E1). In this study, we developed and compared fuzzy[9] and crisp DTs[9] to predict early mortality risk in patients.

Fuzzy decision tree model. The fuzzy DT model applies fuzzy reasoning methods[10] to be able to solve real-world problems more precisely. Therefore, this kind of regression tree uses a fuzzy discriminator function[9] and predicts the degree of membership of each object to the outcome classes.

Survivor and no survivor were 2 outcome classes in this study. We have implemented an entropy error fuzzy decision tree (EEFDT) that applies both entropy[11] and error[9] functions to predict the degree of membership of each patient to each outcome class. Based on the method of Olaru and colleagues,[9] EEFDT implementation consisted of 3 steps: growing, pruning, and refitting steps.

In the growing step, a fuzzy discriminator function found the best fuzzy cut point for each continuous variable. The best predictive variable (among continuous and categorical variables) was selected at each node to reduce the entropy of the dataset at that node (the gain ratio of all continuous and categorical variables was calculated at that node and the variable with the highest gain ratio was selected).[11] This method continued iteratively until termination criteria[9] were met (a growing step was terminated when the cardinality of the local node reached $\leq$10 or the entropy of the local node reached $\leq$0.1 or selecting none of the potential variables resulted in a gain ratio $\geq$0.01). In the pruning step, the irrelevant parts of the grown tree were deleted to increase its interpretability.[9] The nodes were deleted one by one in bottom-up order and the mean absolute error (MAE) of new subtrees was calculated.[9] The best pruned subtree was the smallest one whose MAE was equal to min(MAE) + standard error of MAE that was estimated on the pruning set.[9] In the refitting step, the parameters of the terminal nodes were optimized once more to reduce the error of grown and pruned trees.[9]

Crisp decision tree model. In this study, we also developed an entropy error crisp decision tree (EECDT) to predict early mortality risk. The implementation of this DT was the same as EEFDT. However, EECDT used a crisp discriminator function[9] to find the best cut point for continuous variables in the growing step.

Both DTs were implemented and tested with 3 mutually exclusive subsets. Therefore, the whole dataset was randomly partitioned into 3 subsets: growing, pruning, and testing sets. The sets were stratified, which means that the proportion of dead cases in each dataset was kept around 4% in all sets. The tree was developed with growing, pruned with pruning, refitted with both growing and pruning, and tested with the testing set.

Logistic regression risk model. This model was developed by a combination of the pruning and refitting sets (two-thirds of the dataset) and was tested by the testing set (one-third of the dataset). Based on the full model approach, all predictive variables were retained to develop the model regardless of their statistical significance. A multivariate logistic regression (LR) analysis was then performed to evaluate the independent role of each variable, using probability values of .05 as the threshold for entering variables. In this method, we did not omit the predictive variables that are not statistically significant in the univariate analysis because they could achieve significance when other factors are included in the model.[12] Significant independent variables were entered into the final model and the weight of each variable was obtained from the logistic β coefficient.

Model Performance

The estimates from such a single hold-out cross validation, in which the dataset is partitioned into just 2 mutually exclusive subsets, is somehow biased and depends on the division of the training and testing sets. To get an estimate with lower bias and with potentially better predictive power of our method, we conducted another experiment. We repeated this hold-out cross validation 10 times and the performance of all models was estimated by averaging. Accuracy, AUC, sensitivity, and specificity of the model were reported to evaluate their performance. The method proposed by Vergara and colleagues[13] was used to evaluate the statistical difference between the AUC of the models.

Moreover, EuroSCORE[4] was estimated for patients in the testing set in each hold-out cross validation circle. Accuracy, AUC, sensitivity, and specificity of this model were also calculated by averaging the results.

RESULTS

Patient Characteristics

A total of 1068 adult patients underwent CABG surgery between September 2005 and April 2006 at our institution.