Commentary: Improving the clarity of the crystal ball in cardiac surgery

Timothy M. Guenther, MD,a,b and Lisa M. Brown, MD, MASa

In this issue of the Journal, Ong and colleagues1 demonstrate the use of machine-learning methods to predict mortality after cardiac operations not addressed by the Society of Thoracic Surgeons (STS) adult cardiac surgery risk models. Importantly, the STS risk models perform well for predicting adverse events after coronary artery bypass grafting surgery (CABG), isolated aortic or mitral valve surgery, and combined valve plus CABG procedures.2,3 The authors highlight that 23% of all cardiac surgical procedures nationwide are not represented by these models, and that proportion is even greater at quaternary centers.1 Using preoperative variables included in the STS database, the authors use 4 supervised machine-learning models, 2 linear and 2 nonlinear, to predict operative mortality after cardiac surgery.1 The important distinction between traditional statistical approaches and machine learning is that in machine learning, a model is developed by learning from examples rather than being programmed with rules.4 Moreover, machine-learning methods use greater numbers of mathematical operations than traditional regression methods to better characterize complex relationships between variables and outcomes.2 With regression methods, interactions between variables within the model must be prespecified and tested.

Prediction models must have excellent discrimination, the ability to differentiate those at greater risk of having an event from those at lower risk, and calibration, the ability to accurately predicted absolute risk.6 The C-statistics for operative mortality for CABG, valve, and valve plus CABG using the STS risk models are 0.80, 0.78, and 0.76, respectively.7 The closer a value is to 1.0, the better the model is at correctly classifying outcomes. In the current study, all C-statistics were >0.75, with almost all being ≥0.80 for models trained and tested within the same center.1 In both the STS risk models and the current study, calibration was excellent for most outcomes.1,3

There are several key features of the current study that merit highlighting. First, this study was a collaboration between 3 high-volume academic centers and demonstrated that models developed at one center performed well when tested in a similar center. Second, if the goal is accurate prediction of operative mortality for decision-making or preoperative discussions with patients, it is advantageous to use models developed at the center in which the patient will be undergoing surgery, as the predictions may be more accurate than using models built using nationwide data. In addition, there may be variability in data abstraction from center to center, and using data from a single center to inform machine-learning models is highly likely to decrease that variability. Similar to...
traditional logistic regression methods, machine-learning models are only as good as the data used to build them. Third, machine-learning models can be updated more frequently by continually entering new data, whereas logistic regression models are typically updated at longer intervals. Lastly, machine-learning methods are inherently superior to traditional regression methods at building parsimonious models. It is clinically useful to generate a model using the least number of variables that allows for accurate prediction.

References