EDITORIAL COMMENT

Evaluating the Risk of Coronary Surgery and Percutaneous Coronary Intervention*

William S. Weintraub, MD, FACC
Newark, Delaware

As physicians take their patients through procedures, including coronary artery bypass graft (CABG) surgery and percutaneous coronary intervention (PCI), there will always be concern regarding adverse outcomes. Furthermore, there will be interest in predicting outcomes in advance for each patient or for health care providers. Thus, multiple attempts have been made to estimate risk based on observational outcome databases (1–8). A recent effort to develop models to predict risk of CABG and PCI has been made using the New York State database (9,10). This database offers the distinct advantage of being complete for all patients undergoing coronary revascularization in the state of New York.

In this issue of the Journal, Hannan et al. (9) present a model and validation of the model based on the outcome of CABG in New York State in the year 2002. There were 16,120 patients in the study, with an overall mortality of 2.27%. A logistic model was used to predict mortality based on covariates. This model was then converted into an easy-to-calculate score ranging from 0 to 34. The model included age, gender, hemodynamic state (stable, unstable, shock), ejection fraction, pre-procedure myocardial infarction, chronic obstructive pulmonary disease, extensively calcified ascending aorta, peripheral vascular disease, renal failure requiring dialysis, and previous open heart procedures. The ability of the model to discriminate between those who died and those who did not was good, with a c index of 0.823, where a c index of 0.5 indicates no ability to discriminate and a c index of 1.0 would be perfect discrimination. The calibration of the model, which is the ability to predict risk at varying levels of risk was acceptable with a c index of 0.782. The same risk score could predict non-fatal complications and length of stay.

In a similar analysis in this issue of the Journal, Wu et al. (10) developed a model and validation of the model based on the outcome of PCI in New York State in the year 2002. There were 46,090 procedures performed at 41 hospitals, with an overall mortality of 0.70%. As with CABG, a logistic model was used to predict mortality. The model was converted into an easy-to-calculate score ranging from 0 to 40. The final model included age, gender, hemodynamic state (stable, unstable, shock), ejection fraction, pre-procedure myocardial infarction, peripheral arterial disease, heart failure, renal failure, and left main disease. These variables overlap those in the CABG model, and the definitions are at least similar. The discrimination of the model was excellent, with a c index of 0.886. The calibration was also acceptable with a Hosmer-Lemeshow p value of 0.12. Predicted mortality ranged from 0.05% with a risk score of 0 to 99.36% with a risk score of 40. However, only 0.05% of patients had risk scores of 19 or above, and none higher than 31. The model was validated in 50,046 patients undergoing PCI in 2003, with a c index of 0.905. The same risk score was also found to predict complications and length of stay.

There are many strengths of the risk scores developed in New York State. The scores are easy to calculate, and do not necessarily require a computer or hand-held calculator. The data are also essentially complete for New York, obviating concern over whether these models are from a selected population, as least when applied within New York. When used externally, the risk score can be recalibrated to the mortality in the new location, although with some concern over whether the covariates in the model will be much the same and with the same relative strengths. There are also limitations to this and to all risk models. The model is only as good as the data on which it is based. New York does have an audit system, which limits inaccurate data collection. Nonetheless, there will be concern about the subjectivity of some of the data. A study based on an audit of the American College of Cardiology database revealed that there was greater error or at least greater inconsistency in the collection of subjective data such as acuteness compared to objective data such as age or gender (11).

How can these models be used? The most obvious question is to calculate a risk score for patients before undergoing a procedure. A limitation of using the risk score rather than the full logistic model (technically with the covariant matrix) is that the 95% confidence interval cannot be calculated. It is not difficult to develop a program for hand-held personal digital assistants that will calculate risk explicitly with the confidence interval. This problem relates to determining the precision of a risk score. Whether it would be worthwhile to develop such elegant predictions of risk with estimates of precision is uncertain. This is because a more serious limitation is that there is bias if the risk score is calculated before the procedure as the risk score actually only...
applies to patients who are selected to undergo the procedure. Thus, it would be of uncertain value to calculate a risk score for both PCI and CABG and then decide between them based in part on the risk score. The risk score also cannot take into consideration serious comorbidity in individuals patients not collected in the database used to create the risk score. These are problems related to accuracy. Because of problems related to accurate calculation of risk in any single patient, a risk score is more useful in establishing whether risk is inordinately high or safely low and should be used with considerable skepticism as part of an overall assessment of the clinical situation.

Risk scores or, more properly, the models from which the score is derived, can also be used to risk adjust outcome of different providers, either at the institutional or operator level. This has become especially popular for benchmarking hospitals and physicians. While this is a worthwhile activity, certain caveats should be noted. First, risk adjustment cannot account for unmeasured confounders. Thus, if frailty is not measured, and predicts outcome, then variability in frailty cannot be accounted for. Next, as noted previously, there is subjectivity in some of the measures used in creating models. Thus, acuteness is generally predictive of outcome, and is a subjective measure. If different institutions measure acuteness differently, there will be bias in the application of prediction algorithms across institutions. More simply put, an institution may try to make itself look good by making their patients seem sicker, so-called gaming the system. Next there is a problem with missing data, especially when it is not missing at random (12,13). Typically, ejection fraction is not missing at random as the sickest patients often do not have ventriculograms. If there is a differential collection of ejection fractions across institutions, there will be bias in risk adjustment. Finally, there is uncertainty in the event rate when the measured rate is low. Thus, if an institution performs 400 PCIs a year with two deaths, this would seem to be a quite favorable rate. However, is it really statistically significantly different from an institution performing 400 PCIs with six deaths; what would seem to be a high rate? It is thus difficult to risk adjust when the event rates are so low and the overall number of procedures is modest. This particular problem is more severe for operators, many of whom will have low volume, than for institutions, but the same basic problem applies to both.

Predicting risk-adjusted mortality for hospitals or individual operators, whether for CABG, PCI, or any procedure, is probably most useful as part of a quality improvement process where in a non-punitive fashion the data are fed back to the hospital to facilitate changes in practice that can improve outcome. The Northern New England Study Group has extensive experience with this approach and has documented improvement in outcome, although such improvement may also be due to secular trends (7).

Risk models also have limitations in evaluating quality of care. The models have been best for mortality, but generally have poorer discrimination when specifically developed for complications. There is little experience in developing risk score to evaluate health status or quality of life, which may be what patients value most. Most risk models have also largely been concerned with in-hospital or 30-day mortality rather than long-term outcomes. Risk models also cannot easily be used to evaluate whether procedures are being performed for appropriate reasons, although it is possible with appropriate database design to determine whether procedures are being performed in compliance with guidelines (14). Despite multiple limitations concerning their proper interpretation, risk models clearly are useful. What we seek is medical care that meets the Institutes of Medicine's criteria of quality, that is, care that is safe, effective, patient-centered, timely, efficient, and equitable (15). Risk models mostly fit into this framework by helping to assure that care is safe, but they can also help shape each of the other criteria as well.

Reprint requests and correspondence: Dr. William S. Weintraub, Christiana Care Health Services Stanton-Ogletree Road, Newark, Delaware 19713. E-mail: wweintraub@christianacare.org.

REFERENCES

11. Fiocci FE, Kennett JD, Shaw RE, et al. Accuracy of data is dependent on whether it is subjective or objective: results from the American
College of Cardiology-National Cardiovascular Data Registry Pilot Data Audit Program (abstr). J Am Coll Cardiol 2005;45 Suppl A:33A.