The Eyeball Test
Can the Blind Leading the Blind See Better Than the Statistician?

Karl F. Welke, MD, MS

As physicians, we strive to provide the most appropriate care for our patients and to provide them with accurate assessments of both the risks and benefits they can expect to realize. Unfortunately, as humans, we are notoriously bad at estimating risk. However, as the choices of treatments for cardiac disease expand, accurate assessment of risk becomes increasingly important. A patient with coronary artery disease may have the option of medical therapy, coronary artery bypass grafting, or percutaneous intervention. Transcatheter aortic valve replacement may soon be a viable alternative to surgical aortic valve replacement for patients with aortic valve disease. For patients to be able to make the choice that best suits their goals, they must have accurate information regarding risks and benefits.

Both the risk models and the surgeons’ estimates were better at estimating perioperative mortality than they were at estimating 5-year mortality rates were also considered. Although neither the statistical risk model nor surgeon estimates were perfect, the risk models performed somewhat better at all time points. Both the risk models and the surgeons’ estimates were better at estimating perioperative mortality than they were at estimating late mortality. These findings are informative and may help us to improve our estimates of procedural risk.

When examining the predictive validity of a risk model (or surgeons’ estimates), 2 factors are regularly used: calibration and discrimination. Calibration measures to what degree the model assigns risk to the population, or in a population of 100 patients with an estimated mortality risk of 3%, were there 3 observed deaths? The current article uses the most common measure of calibration, the Hosmer–Lemeshow test for goodness of fit. This test describes the degree of concordance between deciles of observed and expected risk. Discrimination measures the trade-off of specificity and sensitivity at a variety of probability cut points, or how well does the model differentiate between patients who did and did not die? The most common test of discrimination, and the one used in the current article, is the c index which is numerically equivalent to the area under the receiver operating characteristic curve. A model with perfect discrimination would have a c index of 1.0, whereas a model with no ability to discriminate would have a c statistic of 0.50: a coin flip.

The authors found that applied to their population, the risk estimation model for perioperative mortality developed by the VA CICSP was less than ideally calibrated (P<0.0001) and had moderate discrimination (c index, 0.78). The grouped physician risk estimate was also poorly calibrated (P<0.0001) and had slightly worse discrimination (c index, 0.73). There are 2 issues at stake here. One is the ability of surgeons to predict mortality. The other is the ability of a risk model to predict mortality.

Given the amount of data one must process to reach an estimate of mortality risk associated with cardiac surgery, the accuracy of the surgeon’s mortality estimates is commendable. For comparison, The Society of Thoracic Surgeons isolated coronary artery bypass grafting risk model includes 30 variables in 8 categories and has a c index of 0.812. For an individual to retain and process such a volume of information in a limited amount of time is a tall order. Instead, one likely derives an estimate based on knowledge gained during training and gleaned from the medical literature, past personal experience, and key information about the patient in question that differentiates them from the norm. During the period of the study, 2 additional events occurred that may have influenced surgeons’ mortality risk estimates. Risk calculators became more readily available, and the attention to operative outcomes both for quality improvement and public reporting purposes increased. Although not examined in the present study, surgeons’ estimates may have improved over time attributable to increased availability of and attention to mortality statistics.

A risk model works best when estimating risk for the population from which it was derived. The more divergent the population of interest is from the population from which the model was derived, the less valid the model. For example, the VA CICSP model was derived from a population of primarily men who underwent cardiac surgery in VA hospitals. The model would be less likely to estimate accurately the risk of women undergoing cardiac surgery in non-VA hospitals. Similarly, The Society of Thoracic Surgeons risk models and the EuroSCORE, although carefully constructed and rigorously evaluated, may not assess the risk of the veteran population as well as the VA CICSP model because they were derived from different patient populations undergoing surgery in different
hospital environments. In addition, because more patients who undergo cardiac surgery are of relatively low rather than relatively high mortality risk, model estimates are more accurate for lowest-risk rather than highest-risk patients. As systems and knowledge change over time, a model is also best suited to the year or years in which the included population underwent surgery. Because cardiac surgery mortality rates have in general improved with time, a risk model from a previous time period will overestimate the current mortality risk associated with surgery. Therefore, it is necessary to periodically recalibrate risk models as practices and outcomes change.

A risk model may also be limited in its ability to accurately predict an outcome if it omits variables associated with the outcome. Despite the extensive lists of variables included in clinical databases, significant gaps persist. Patient-level factors, such as patient report health status, have been shown to be independently associated with mortality after cardiac surgery yet are not commonly included in clinical databases. Frailty and nutritional status too may impact mortality but are not routinely collected. Risk models tend to focus on patient-specific factors and neglect broader surgeon and center variables. Surgeon experience and center-level factors, such as operative volume and intensive care unit staff expertise and availability, also impact outcome but are not routinely included in surgical risk models. When physicians assess patients and assess their mortality risk, they may be including some or all of these additional factors in their subjective calculation.

Risk models for estimating operative risk are now readily available. The Society of Thoracic Surgeons risk model for cardiac surgery and the EuroSCORE are available online. With such tools easily accessible, the need for physicians to estimate risk based solely on past experience alone is unnecessary. Instead, personalized risk estimates can be calculated for individual patients. Yet, one must exhibit caution when using cardiac surgery risk models to predict individual patient risk. These models were developed for estimating the risks of large populations rather than individual patients. Most such models are well calibrated but produce estimates that are extreme in the low- and high-risk groups of patients. In addition, the calibration of most of these risk models is at best moderate. A model may accurately predict that 3 of 100 patients will have a perioperative mortality, but cannot identify which 3 patients will die. Because the predictive ability of these models is not perfect, there is still a need for clinicians to interpret the model-derived estimates and perhaps alter them for patients less well represented in the population used for development of the model. This may include high-risk patients or those with unusual conditions. Short of these special circumstances, however, physicians can provide patients with fairly accurate assessments of their perioperative mortality risks without the need for guesswork. By adding knowledge of local surgeon and center-specific factors to fine-tune the risk derived from a statistical risk model, the estimates may be even more accurate.

Long-term risk is more difficult to predict than operative mortality. Fewer studies have looked at this topic, and there are few risk models designed for this purpose. The further out a patient is from their operation, the more the risk of mortality is influenced by factors other than those included in perioperative risk models. The patient may develop health problems not present at the time of surgery, or conditions present before surgery may subsequently improve. Therefore, the predictive value of such models, including the one used in the present study would be expected to decline over time. Similarly, most surgeons are more knowledgeable about perioperative rather than long-term risk, and it is not surprising that the accuracy of their estimates was worse for the 1- and 5-year time points.

As physicians, we are obligated to provide our patients with the most accurate information possible regarding their conditions and the risks they present. To do so, we need to improve the validity of statistical risk models by expanding them to include additional predictive variables and keep them up to date. Risk modeling must continue to expand beyond mortality. At a perioperative mortality rate of 3.3%, 96.7% of patients survive surgery and are likely interested in outcomes in addition to mortality. Many are undergoing cardiac surgery to improve their quality of life, yet our ability to predict their odds of doing so are restricted. We should not only accept our limitations in independent, subjective risk prediction but also understand that our assessment of the patient and our knowledge of the care environment added to that derived from a statistical risk model may result in the most accurate assessment of an individual patient’s risk.

Disclosures
None.

References

Key Words: Editorials • models, statistical • thoracic surgery
The Eyeball Test: Can the Blind Leading the Blind See Better Than the Statistician?
Karl F. Welke

Circ Cardiovasc Qual Outcomes. published online January 14, 2014;
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/early/2014/01/13/CIRCOUTCOMES.113.000760.citation

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/