Ranking of Surgical Performance

To the Editor:

Ranking of surgical performance is very difficult. Recently, operative mortality after CABG was studied with a focus on the practical application of risk indexes.1 I have 3 methodological comments.

In the accompanying editorial,2 it was suggested that a common problem in statistical models is “to predict a higher risk of death than is observed for the patients with the least severe illness and a lower risk of death than is observed for the most ill patients.” This is contradicted by empirical studies, including Ivanov’s.1 Likely, the display in Figure 1 has been deceptive (predictions on the y axis, observations on the x axis). Advanced statistical research has addressed the phenomenon of too extreme predictions in independent data, ie, that low predictions are too low and high predictions too high. A practical solution is to “shrink” the regression coefficients, such that predictions for future patients are better calibrated.3

Second, an additive score was based on the rounded odds ratios in the logistic regression model.1 Although this score is probably meant only for illustrative purposes, it should have been based on the regression coefficients (ie, the natural logarithm of the odds ratios). The differences in scoring are minor in the models considered, except for “LV [left ventricular] grade=4,” which gets a weight of 13 or 14 in the “recalibrated” and “remodeled” models, respectively. These weights should have been around 7. For the recalibrated model, a score of 13 can be obtained for a patient with LV grade 4 and no other risk factors, or with LV grade 2 or 3 and presence of risk factors, while the corresponding correct probabilities are 5% or around 25%, respectively. This makes Table 3 hard to interpret.

Finally, the editorial rightly stressed the importance of sample size in ranking.2 Statistical power is determined by the 169 operative deaths, not by the impressive number of 7500 patients registered. The low number of events makes “remodeling” an unstable strategy. More importantly, the ranking between surgeons will to a large extent be determined by chance, since the average number of events was around 12 for the 14 surgeons. The differences between surgeons are probably smaller than observed. Especially for surgeons who have performed relatively few procedures, the evidence is limited against the null hypothesis of similar operative skills between surgeons. Furthermore, the uncertainty at the individual level results in substantial uncertainty in the ranking as a whole.4,5 This argues for a cautious interpretation of rankings, independent from the specific type of risk adjustment.

Ewout W. Steyerberg, PhD
Department of Public Health
Erasmus University Rotterdam, The Netherlands


Response

We appreciate the opportunity to respond to the letter written by Dr Steyerberg and will attempt to address each of his 3 major concerns.

The purpose of our article1 was to evaluate 3 different modeling strategies for the evaluation of risk-adjusted operative mortality after CABG. The concern about overestimation and underestimation of the predicted probability of outcomes has been evaluated previously by Spiegelhalter,2,3 whose aim was to determine whether “shrinkage” of predictions toward an origin would achieve more accurate probabilistic predictions. Spiegelhalter compared estimates calculated from a logistic model with those of a shrunken model3 and found that although the shrunken model resulted in variables that had less of an impact on the final calculation of probability, there was very little actual difference between the 2 models. Indeed, the C statistic (analogous to the area under the receiver-operator characteristic curve for a binomial response variable) was identical in both models, and Brier scores, which evaluated precision at the individual patient level, were 0.186 and 0.182, respectively. We agree with the argument put forth by DeLong and colleagues4 that a higher level of sophistication may be achieved by calculating “shrunken” estimates but that the results are probably more philosophical than practical. Shrinkage may provide for more precise estimates, especially at the tails; however, we believe Figure 1 is not deceptive but rather a fair representation of the impact of the different modeling strategies.

The purpose of using rounded odds ratios to calculate a total risk score was to construct a simple, clinically accessible method to categorize patients into relative risk groups. Additionally, using the odds ratios to compute total risk scores allowed us to compare predicted versus observed probabilities using general linear regression and analysis of covariance without having to further transform the values on the y axis and therefore making this evaluation more transparent to the average clinician. We would like to take this opportunity to correct our typographical error in Table 2. The weight for LV grade 4 should read 12 and not 14. Regression coefficients and not total risk scores were used to calculate predicted probabilities. Observed operative mortality for the 220 patients having LV grade 4 was 12.3%, which is well within the confidence interval for the predicted probability of operative mortality for this group of 11.9±0.7%. These results suggest that even for high-risk patients, the remodeled strategy was reasonably precise.

DeLong and colleagues5 evaluated 8 different modeling strategies in 3654 patients in 28 centers with an operative mortality rate of 5.6% (n=205) and found that the methods (including shrinkage) drew similar conclusions regarding outlier status. The purpose of Figure 2 and Table 5 was to depict the impact of the 3 different modeling strategies on the relative ranking of providers and not the specific rankings themselves. In this regard, we believe we had sufficient sample size and outcome events to
compare the modeling strategies within the same population of patients.

In conclusion, we thank Dr Steyerberg for his insightful comments and draw on the observations of Dr Krumholz that more research is needed into the validity of risk-adjustment methods.

Joan Ivanov, MSc
Division of Cardiovascular Surgery
Toronto Hospital
Toronto, Ontario, Canada

Jack V. Tu, MD, PhD
C. David Naylor, MD, DPhil
The Institute for Clinical Evaluative Sciences
North York, Ontario, Canada

Response
I appreciate Dr Steyerberg’s interest in my editorial. In his letter, Dr Steyerberg disagrees with my comment about the calibration of risk models. I was suggesting that it is difficult for predictive models to capture the true risk of patients at the extremes of risk. For example, a dichotomous variable for shock does not capture the grades of severity among patients with hypoperfusion. Because of the coarseness of many variables, these models often cannot capture the severity of the most ill patients, for whom surgery is a perilous but potentially life-saving treatment. In addition, Ivanov’s results do not contradict my point, as suggested by Dr Steyerberg, since the article did not report any patient-level comparisons of predicted and observed mortality rates.

Harlan M. Krumholz, MD
Yale School of Medicine
New Haven, Conn

Ranking of Surgical Performance
Ewout W. Steyerberg

_Circulation_. 2000;102:e61-e62
doi: 10.1161/01.CIR.102.9.e61
_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2000 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circ.ahajournals.org/content/102/9/e61

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ 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_ is online at:
http://circ.ahajournals.org//subscriptions/