Operative Mortality and Procedure Volume as Predictors of Subsequent Hospital Performance

John D. Birkmeyer, MD, Justin B. Dimick, MD, MPH,*† and Douglas O. Staiger, PhD‡*

Context: Despite growing interest in evidence-based hospital referral for selected surgical procedures, there remains considerable debate about which measures should be used to identify high-quality providers.

Objectives: To assess the usefulness of historical mortality rates and procedure volume as predictors of subsequent hospital performance with different procedures.

Design, Setting, and Participants: Using data from the national Medicare population, we identified all U.S. hospitals performing one of 4 high-risk procedures between 1994 and 1997. Hospitals were ranked and grouped into quintiles according to 1) operative mortality (adjusted for patient characteristics) and 2) procedure volume.

Main Outcome Measures: Risk-adjusted operative mortality in 1998 to 1999.

Results: Although historical mortality and volume both predicted subsequent hospital performance, the predictive value of each varied by procedure. For coronary artery bypass graft surgery, mortality rates in 1998 to 1999 differed by 3.3% across quintiles of historical mortality (3.6% to 6.9%, best to worst quintile, respectively), but only by 1.0% across volume quintiles (4.8% to 5.8%). In contrast, for esophagectomy, mortality rates in 1998 to 1999 differed by 12.5% across volume quintiles (7.5% to 20.0%, best to worst quintile, respectively), but only by 1.5% across quintiles of historical mortality (11.4% to 12.9%). Historical mortality and procedure volume had comparable value as predictors of subsequent performance for pancreatic resection and elective abdominal aortic aneurysm repair. Our findings were similar when we repeated the analysis using data from later years.

Conclusions: Historical measures of operative mortality or procedure volume identify hospitals likely to have better outcomes in the future. The optimal measure for selecting high-quality providers depends on the procedure.

(*Ann Surg* 2006;243: 411–417)

In light of wide variation in surgical performance with many procedures, efforts are currently underway to direct patients toward the highest-quality hospitals. The Leapfrog Group, a large coalition of healthcare purchasers, has implemented standards for "evidence-based hospital referral" for 5 high-risk procedures.[1] Other efforts are aimed at disseminating provider-specific information about surgical performance with the hopes that patients will select higher-quality hospitals. In addition to information provided by proprietary sources (eg, Healthgrades.com), a growing number of states are reporting hospital-specific measures of surgical quality. For example, the Texas Health Care Information Council recently began public reporting of both procedure volume and mortality rates for several surgical procedures.[2]

Despite this growing interest in assessing surgical quality, there remains controversy about how best to identify high-quality hospitals for individual procedures. Hospital procedure volume is currently among the most widely used quality indicators. There remains little doubt that volume is inversely related to operative mortality with many procedures.[3–5] Nonetheless, critics decry volume as a crude surrogate for quality and a poor predictor of individual hospital performance.[6–8] Instead, many think that surgical quality is best judged by direct outcome measures, including operative mortality. For many procedures, however, hospital mortality rates may be hampered by sample size problems and thus may be too imprecise to meaningfully reflect quality of care.[6,9]

Studies assessing the value of different quality indicators have generally focused on their relative abilities to describe hospital performance in a previous time period.[10,11] However, since most scorecards are intended primarily to improve current decision making by patients or purchasers, surgical quality indicators are perhaps better judged by how well historical measures predict future hospital performance.[12,13] In this context, we used data from the national Medicare population to compare the relative usefulness of 2 measures (hospital volume and operative mortality) as predictors of subsequent surgical mortality. In essence, this analysis considers a hypothetical Medicare patient deciding

From the *Michigan Surgical Collaborative for Outcomes Research and Evaluation (M-SCORE), Department of Surgery, University of Michigan, Ann Arbor, MI; †VA Outcomes Group, VA Medical Center, White River Junction, VT; and ‡Department of Economics and the Center for the Evaluative Clinical Sciences, Dartmouth College, Hanover, NH.

Supported by the Agency for Healthcare Research & Quality (R01 HS10141-01), which had no role in the design or conduct of the study; collection management, analysis and interpretation of the data; and preparation, review, and approval of the manuscript. Supported by NIH/NIA grant P01 AG19783-01. Dr. Birkmeyer is a consultant for the Leapfrog Group and chairs its Expert Panel on evidence-based hospital referral.

The views expressed herein do not necessarily represent the views of Center for Medicare and Medicaid Services or the United States Government.

Reprints: John D. Birkmeyer, MD, 2920 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109. E-mail: jbirkmey@med.umich.edu.

ISSN: 0003-4932/06/24303-0411
DOI: 10.1097/01.sla.0000201800.45264.51

where to have 1 of 4 high-risk procedures in 1998 or 1999, based on hospital "scorecards" created using historical (1994–1997) data. Should the patient choose a high-volume hospital, or one with low operative mortality?

METHODS

Subjects and Databases

We used 100% national analytic files from the Center for Medicare and Medicaid Services for years 1994 through 2001. MEDPAR files, which contain hospital discharge abstracts for all fee-for-service acute care hospitalizations of all U.S. Medicare recipients, were used to create our main analysis datasets; the denominator file was used to assess patient vital status at 30 days postsurgery.

Using appropriate procedure codes from the *International Classification of Diseases*, version 9 (ICD-9 codes), we identified all patients 65 to 99 years of age undergoing 1 of 4 surgical procedures selected by the Leapfrog Group for its evidence-based hospital referral initiative: coronary artery bypass grafting (CABG), elective abdominal aortic aneurysm (AAA) repair, pancreatic resection, and esophagectomy.[14,15] We made several restrictions to align our analyses with the surgical populations targeted by the Leapfrog Group and to avoid skewing our results by a small number of higher-risk patients. Thus, patients undergoing valve replacement were excluded from the CABG cohort (12%). The AAA repair cohort excluded patients with ruptured or thoracoabdominal aneurysms (33%). Finally, patients with noncancer diagnoses were excluded from the pancreatic and esophageal resection cohorts (15%).

Hospital Volume and Prior Mortality

We calculated average procedure volumes and risk-adjusted operative mortality rates for each U.S. hospital performing at least 1 of the 4 procedures between 1994 and 1997. For each procedure, we determined the total number performed by each hospital in Medicare patients over the 4-year period. Medicare volumes were converted to estimates of total (all-payer) volume at each hospital using procedure-specific multipliers derived from the 1997 Nationwide Inpatient Sample, as described previously.[5] We then ranked hospitals by their average volumes and applied hospital volume cut-points that most closely sorted the patient sample into 5 evenly sized groups (quintiles). Volume thresholds used for each procedure are shown in Table 1.

For each procedure, we also determined risk-adjusted operative mortality rates for each hospital. Operative mortality was defined by death occurring before hospital discharge or within 30 days of surgery. Risk adjustment was performed using methods previously described.[5] Variables in our risk adjustment models included age group (5-year intervals), sex, race (black, nonblack), admission acuity (elective, urgent/emergent), and mean Social Security income (zip code level). Comorbidities were identified by their appropriate ICD-9 codes and aggregated into Charlson scores with published weights.[16,17] To develop and evaluate our risk-adjustment models, we determined the discrimination and calibration of the logistic regression model for each operation. Measures of discrimination showed moderate predictive ability (C-statistics ranged from 0.68 to 0.70 across operations). Hosmer-Lemeshow tests did not reject the models ($P > 0.1$ for all operations, indicating good calibration).

Hospitals were then ranked and placed into quintiles of historical mortality. Because mortality rates at individual hospitals are often "noisy" when small numbers of cases are performed, we based our rankings on t-statistics. The t-statistic is the difference between a hospital's observed and expected mortality rate, divided by the standard error of the expected mortality rate. In this way, this measure accounts for both the hospital's relative performance and its sample size. Thus, our mortality rankings were based on the statistical likelihood that a hospital's mortality rate was worse (or better) than expected (Table 1), the approach taken by most public reporting systems for surgical mortality. The net effect of this approach was to dampen extreme mortality rates observed at hospitals with very low caseloads, moving them toward the middle of hospital rankings.

We used the risk-adjustment models from these analyses to calculate the expected mortality rates in each of the volume and historical mortality quintiles. Based on each patient's set of baseline characteristics, the logistic regression equation provides an estimate of the predicted probability of death. To calculate the expected rates for each quintile, we summed the predicted probabilities of death for all patients undergoing surgery at hospitals in that quintile.

Prediction of Subsequent Performance

Subsequent mortality was determined using 1998–1999 Medicare data. To assess how well each measure (historical volume and mortality) predicted subsequent mortality, we created a separate random-effects logistic regression model for each operation. We used random-effects models because they account for nonindependence of observations within hospitals (ie, clustering). The patient was used as the unit of analysis; operative mortality was the dependent variable. Quintiles of historical volume (or mortality) were entered as independent variables, along with patient characteristics used for risk adjustment. To describe the predictive ability of each measure, we calculated risk-adjusted mortality rates for each of the volume and historical mortality quintiles. We also calculated the odds ratios of mortality at hospitals in the "worst" quintile, relative to those in the "best" quintile for each procedure. In these analyses, we tested the statistical significance of the trends across quintiles by entering the categorical quintile variable (rather than dummy variables) into the regression model.

Using a second set of random-effects logistic regression models, we also estimated the proportion of variation in subsequent mortality explained by each of the 2 historical measures.[12] In such models, the random effect reflects unexplained hospital-level variation in (1998–1999) mortality rates. For each procedure, the proportion of variation explained by each measure was assessed by the percent reduction in the standard deviation of the random effect as either historical mortality or volume (assessed as continuous variables) was added to the model. For esophagectomy, there were too few cases per hospital to generate stable estimates. Thus, these analyses were restricted to the 3 other procedures.

TABLE 1. Actual (Risk-Adjusted) and Expected Mortality Rates for Volume and Mortality Quintiles in Medicare Patients Undergoing Four Procedures, 1994–1997

	Quintile of Historical Mortality or Volume (1994–1997)				
	1 (Worst)	**2**	**3**	**4**	**5 (Best)**
Coronary artery bypass grafting (Medicare, n = 612,373)					
Hospitals ranked by risk-adjusted mortality					
Actual mortality (%)	8.2	6.2	5.3	4.3	3.4
Expected mortality (%)	5.3	5.4	5.4	5.4	5.6
Hospitals ranked by procedure volume					
Average annual volumes	<217	217–342	343–512	513–767	>767
Actual mortality (%)	6.0	5.7	5.3	5.1	5.0
Expected mortality (%)	5.5	5.4	5.4	5.5	5.5
Elective abdominal aortic aneurysm repair (Medicare, n = 95,295)					
Hospitals ranked by risk-adjusted mortality					
Actual mortality (%)	12.3	7.2	4.4	2.7	2.1
Expected mortality (%)	5.8	5.7	5.7	5.7	5.9
Hospitals ranked by procedure volume					
Average annual volumes	<11.8	11.8–21.5	21.6–35.0	35.1–57.3	>57.3
Actual mortality (%)	7.8	5.9	5.3	5.3	4.3
Expected mortality (%)	6.0	5.7	5.7	5.7	5.6
Esophageal cancer resection (Medicare, n = 4349)					
Hospitals ranked by risk-adjusted mortality					
Actual mortality (%)	53.1	18.6	2.9	1.9	2.1
Expected mortality (%)	16.6	16.2	14.5	16.1	15.7
Hospitals ranked by procedure volume					
Average annual volumes	<1.3	1.3–2.0	2.1–3.0	3.1–7.3	>7.3
Actual mortality (%)	21.8	17.1	16.9	13.3	8.1
Expected mortality (%)	16.8	16.3	15.9	15.3	14.2
Pancreatic cancer resection (Medicare, n = 6896)					
Hospitals ranked by risk-adjusted mortality					
Actual mortality (%)	40.1	12.0	1.7	1.4	1.8
Expected mortality (%)	11.9	11.1	11.8	11.8	10.6
Hospitals ranked by procedure volume					
Average annual volumes	<1.8	1.8–2.5	2.6–5.0	5.1–13.5	>13.5
Actual mortality (%)	17.3	15.5	11.0	8.0	4.4
Expected mortality (%)	12.3	11.5	11.6	11.3	10.1

Each quintile contains approximately 20% of patients undergoing that procedure.

Sensitivity Analysis

In sensitivity analysis, we used data from different time periods to assess both historical performance and subsequent mortality. First, to test the robustness of our findings, we repeated the same analysis but moved it forward by 2 years. Thus, historical mortality and volume were assessed using 1996–1999 data, subsequent mortality from 2000–2001 data. Second, we sought to determine how well the predictive value of historical measures held up over time, particularly relevant given real world lags in the availability of data for assessing past performance. In this analysis, we used 1994–1997 data to determine historical mortality and volume but 2000–2001 data to assess subsequent mortality.

RESULTS

Historical Mortality and Procedure Volume

Between 1994 and 1997, approximately 719,000 Medicare patients underwent 1 of 4 procedures, the large majority undergoing CABG. As shown in Table 1, average mortality rates varied widely across hospital quintiles based on both mortality and procedure volume. Differences in risk-adjusted mortality between "best" and "worst" were larger for mortality quintiles than for volume quintiles for all 4 procedures. These differences were particularly large for the 2 high-risk but uncommon procedures. For example, with esophagectomy, hospitals in the best mortality quintile had an average, risk-adjusted mortality rate of 2.1%, versus 53.1% for hospitals in the worst quintile. In contrast to actual mortality rates, expected mortality rates varied little across either mortality or volume quintiles, indicating few measurable differences in patient case mix among the hospital groups.

Prediction of Subsequent Mortality

We then determined the extent to which mortality rates and procedure volumes from 1994–1997 predicted mortality during the subsequent 2-year period (1998–1999) (Table 2; Fig. 1). Historical mortality predicted subsequent mortality

TABLE 2. Prediction of Subsequent Hospital Mortality Rates (1998–1999) by Historical Mortality and Procedure Volumes (1994–1997), Expressed in Odds Ratios of Mortality

	Predictors of Subsequent Mortality (1998–1999)			
	Historical Mortality (1994–1997)		Procedure Volume (1994–1997)	
	Odds of Subsequent Mortality, Worst vs. Best Quintile (95% CI)	Proportion of Variation Explained (%)	Odds of Subsequent Mortality, Worst vs. Best Quintile (95% CI)	Proportion of Variation Explained (%)
Coronary artery bypass grafting	2.00 (1.88–2.10)	54	1.24 (1.18–1.31)	9
Abdominal aortic aneurysm repair	1.84 (1.61–2.10)	35	1.51 (1.32–1.71)	26
Pancreatic cancer resection	5.10 (3.13–8.32)	41	5.84 (3.59–9.48)	50
Esophageal cancer resection	1.18 (0.75–1.89)	*	3.09 (1.90–5.02)	*

*The number of cases within individual hospitals was too small for esophageal resection to allow a stable estimate of the proportion of variation explained.

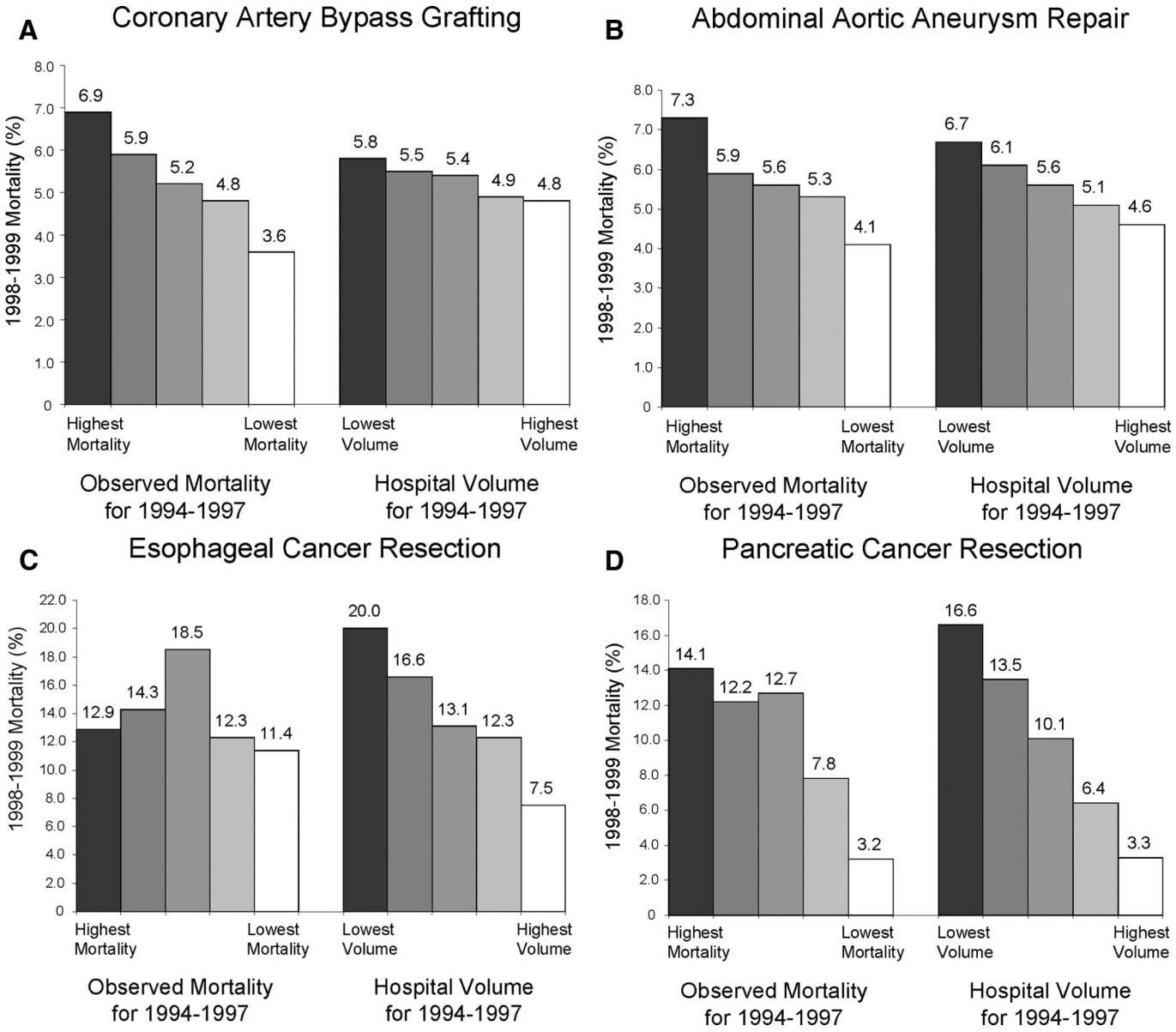

FIGURE 1. A–D, Hospital mortality rates in 1998–1999, according to quintiles of historical mortality and procedure volume (1994–1997). Both historical and subsequent mortality rates are adjusted for patient characteristics.

for coronary artery bypass, elective AAA repair, and pancreatic resection ($P < 0.001$ for each procedure), but not esophagectomy. Procedure volume predicted subsequent mortality for all 4 procedures ($P < 0.001$ for each procedure).

The relative ability of the 2 measures to predict subsequent outcomes varied according to procedure. For coronary artery bypass surgery, historical mortality was a stronger predictor of subsequent mortality than procedure volume. Subsequent mortality rates differed by 3.3% across quintiles of historical mortality (3.6%–6.9%, best to worst quintile, respectively), but only by 1.0% across volume quintiles (4.8%–5.8%). (Fig. 1) In contrast, procedure volume predicted subsequent mortality considerably better than historical mortality for esophagectomy. Subsequent mortality rates differed by 12.5% by volume quintile (7.5%–20.0%, best to worst quintile, respectively), but only by 1.5% across quintiles of historical mortality (11.4%–12.9%). In terms of absolute differences in subsequent mortality rates, historical mortality and procedure volume had comparable value as predictors of performance with pancreatic resection and elective abdominal aortic aneurysm repair.

Similarly, the 2 historical measures of performance differed in their ability to explain hospital-level variation in subsequent mortality (Table 2). For CABG, historical mortality explained 54% of the variation in subsequent mortality, while hospital volume explained only 9%. Historical mortality and volume explained similar amounts of variation in subsequent mortality with elective AAA repair (35% versus 26%, respectively) and pancreatic resection (41% versus 50%).

Sensitivity Analysis

To test the robustness of our findings, we first repeated the same analysis after moving forward the time period of the analysis by 2 years (1996–1999 historical measures versus 2000–2001 mortality). For all 4 procedures, this analysis yielded very similar results to our baseline analysis using data from 2 years earlier (Table 3). Second, to explore the implications of delays in data availability, we used 1994–1997 data for historical measures and 2000–2001 data for subsequent mortality (Table 3). Inserting this 2-year "time-lag" had negligible effect on the predictive value of historical mortality and procedure volume for CABG, elective AAA repair, and pancreatic resection. For esophagectomy, adding a time lag did not diminish the predictive ability of historical volume. However, historical mortality remained a poor predictor of subsequent performance.

DISCUSSION

Most studies assessing the value of different quality indicators have focused on their relative abilities to describe concurrent hospital performance. Thus, volume-outcome studies have generally examined to what extent hospital procedure volumes are associated with mortality during the same time period.[3–5] Studies assessing hospital performance with operative mortality rates (with either clinical or administrative data) have tended to focus on the discriminative ability of risk adjustment models (ie, C statistics) derived from a single, retrospective dataset.[10,11] While such studies may be a good place to start, patients and purchasers are primarily interested in knowing which providers are likely to have the best outcomes now, not several years ago. In this "scorecard" context, quality indicators are perhaps best judged by how well historical measures predict subsequent hospital performance.

Few studies have assessed the prognostic value of historical performance measures in surgery. In our study, historical mortality rates strongly predicted subsequent mortality for CABG, elective AAA repair, and pancreatic resection, even after accounting for potential lags in data availability. Our findings echo results from a recent study of mortality in neonatal intensive care units.[12] Similarly, in an earlier study of coronary artery bypass in California, Luft and Romano noted that hospitals identified as high outliers tended to have significantly higher than expected mortality rates 2 years later.[18] To our knowledge, ours is the first study to describe the extent to which historical procedure volume predicts subsequent mortality. For all 4 procedures, prior procedure volumes were as strongly related to subsequent mortality rates as they were to concurrent mortality. Thus, although the importance of procedure volume varies widely by procedure, the prognostic value of hospital volume is remarkably stable over time.

In terms of prognosis, the optimal quality indicator depends on the procedure. With CABG, historical mortality was much better than volume in predicting subsequent mortality. This should not be surprising. Caseloads at individual hospitals are relatively high with this procedure, allowing for relatively precise estimates of provider-specific mortality. Moreover, hospital procedure volume tends to be weakly associated with mortality with this procedure compared with other higher-risk procedures.[3–5] In contrast, with esophagectomy, procedure volume predicts subsequent hospital performance much more consistently than historical mortality. This high-risk operation has a well-known and particularly strong volume-outcome association. Moreover, because it is performed very infrequently at most hospitals, historical mortality rates tend to be very imprecise and vary considerably from year to year. This no doubt explains the inconsistent relationships between historical mortality and subsequent mortality observed in this study.

Of course, neither historical mortality nor procedure volume was a perfect predictor of future performance at individual hospitals. The 2 measures explained no more than half of hospital-level variation in subsequent mortality with any procedure. Nonetheless, our results suggest that for many procedures patients would reduce their operative mortality risks on average by selecting a hospital in the "best" category for either historical mortality or volume.

It is important to acknowledge several limitations. First, we focused on only 4 procedures: those targeted by Leapfrog Group for evidence-based hospital referral. Thus, our results do not provide guidance on optimal quality indicators for other procedures. Second, our study focused on Medicare patients, who account for just over half of all U.S. patients undergoing the 4 procedures and a larger proportion of patients dying after surgery. Our reliance on Medicare data is a limitation in 2 respects. First, in using Medicare volumes to

TABLE 3. Sensitivity Analysis Showing Relationships Between Historical Mortality and Procedure Volume and Subsequent Hospital Mortality Rates When Assessed Using Data From Different Time Periods

	Subsequent Mortality (%), by Quintile of Historical Mortality or Volume				
	1 (Worst)	2	3	4	5 (Best)
Coronary artery graft bypass					
Historical mortality					
1994–1997 ranking vs. 1998–1999 mortality	6.9	5.9	5.2	4.8	3.6
1996–1999 ranking vs. 2000–2001 mortality	6.4	5.7	4.8	4.2	3.8
1994–1997 ranking vs. 2000–2001 mortality	6.3	5.4	4.9	4.5	3.8
Historical volume					
1994–1997 ranking vs. 1998–1999 mortality	5.8	5.5	5.4	4.9	4.8
1996–1999 ranking vs. 2000–2001 mortality	5.4	5.3	4.8	4.7	4.6
1994–1997 ranking vs. 2000–2001 mortality	5.5	5.2	4.8	4.8	4.6
Elective abdominal aortic aneurysm repair					
Historical mortality					
1994–1997 ranking vs. 1998–1999 mortality	7.3	5.9	5.6	5.3	4.1
1996–1999 ranking vs. 2000–2001 mortality	7.3	6.3	5.9	5.6	4.8
1994–1997 ranking vs. 2000–2001 mortality	7.4	6.3	6.2	5.5	4.5
Historical volume					
1994–1997 ranking vs. 1998–1999 mortality	6.7	6.1	5.6	5.1	4.6
1996–1999 ranking vs. 2000–2001 mortality	7.5	5.9	5.5	5.6	5.3
1994–1997 ranking vs. 2000–2001 mortality	7.3	6.0	5.6	5.5	5.3
Esophageal cancer resection					
Historical mortality					
1994–1997 ranking vs. 1998–1999 mortality	12.9	14.3	18.5	12.3	11.4
1996–1999 ranking vs. 2000–2001 mortality	16.7	12.9	15.1	13.0	9.4
1994–1997 ranking vs. 2000–2001 mortality	15.9	11.4	20.6	12.0	8.0
Historical volume					
1994–1997 ranking vs. 1998–1999 mortality	20.0	16.6	13.1	12.3	7.5
1996–1999 ranking vs. 2000–2001 mortality	18.9	14.1	12.7	12.1	7.5
1994–1997 ranking vs. 2000–2001 mortality	18.9	14.6	12.4	11.5	8.1
Pancreatic cancer resection					
Historical mortality					
1994–1997 ranking vs. 1998–1999 mortality	14.1	12.2	12.7	7.8	3.2
1996–1999 ranking vs. 2000–2001 mortality	14.3	10.8	11.5	8.4	3.5
1994–1997 ranking vs. 2000–2001 mortality	15.3	12.1	11.7	7.3	3.5
Historical volume					
1994–1997 ranking vs. 1998–1999 mortality	16.6	13.5	10.1	6.4	3.3
1996–1999 ranking vs. 2000–2001 mortality	14.1	13.6	11.3	5.4	3.5
1994–1997 ranking vs. 2000–2001 mortality	14.5	14.4	11.6	5.8	3.2

All trends were statistically significant ($P < 0.01$), with the exception of historical mortality for esophageal resection.

estimate total hospital volumes, we may have misclassified the true volume status of some hospitals, which would tend to bias our analysis toward underestimating the prognostic value of this measure. However, we suspect the magnitude of this potential volume misclassification to be small. Our previous analyses, based on data from the 1997 Nationwide Inpatient Sample, have suggested very high correlations between Medicare-only hospital procedure volumes and all-payer volumes, with correlations between 0.90 and 0.98 for most procedures. Second, and more importantly, our reliance on Medicare data alone would tend to increase the imprecision associated with historical mortality rates (due to smaller sample sizes) and thus reduce their predictive value, particularly for infrequent procedures. To augment the precision of our estimates, we used 4 years of data in determining historical mortality rates, a longer time interval than used in conventional hospital report cards. Nonetheless, historical mortality measures from all-payer databases (eg, from the Healthcare Utilization Project) may be better for predicting future surgical outcomes.

Finally, many would criticize our use of administrative data with respect to risk adjustment. Administrative data are no doubt flawed in their ability to capture patient illness severity relative to clinical data.[19] However, risk adjustment

is only important if illness severity varies across hospitals. With the possible exception of coronary artery bypass,[10,11] such variation in case mix among patients undergoing other procedures has not been established. If case mix did vary but not systematically (ie, some hospitals have sicker patients than average in some years, healthier in others), inability to fully capture illness severity with administrative data would lead to underestimation of the true correlation between historical and future mortality. However, if case mix varied systematically across hospitals (ie, some hospitals consistently treat sicker patients), using administrative data would tend to overestimate the true correlation between past and future performance. Although it is not clear whether our results would have differed if we had access to detailed clinical information for better risk adjustment, this question may be moot from a practical perspective. With the exception of cardiac surgery, clinical data for determining risk-adjusted mortality rates with other procedures are currently not on the horizon.

Whether ultimately based on clinical or administrative data, efforts to improve the prediction of hospital performance will need to go beyond individual measures. Composite measures that simultaneously account for multiple quality indicators appear promising in initial applications.[13] For surgery, optimal measures would need to incorporate numerous structural variables not considered in this analysis. Potential candidates include hospital experience with other procedures, intensive care unit staffing, nurse staffing levels, and surgeon volume and specialty training.[20–24] Optimal measures could incorporate data on processes of care related to lower mortality, eg, use of perioperative beta-blockade in high-risk patients.[25] Finally, optimal measures would account for not only hospital mortality with the procedure of interest, but mortality with other, related procedures.[26]

Until such tools are available, purchasers and policy makers should be thoughtful in selecting quality indicators appropriate for different procedures. Currently, a wide array of surgical quality measures are being used to identify high-quality providers, in efforts ranging from selective referral initiatives by purchasers to public reporting systems aimed at informing patients. However, the usefulness of many of these measures has not been established. Quality indicators under consideration for surgical scorecards should be assessed empirically, with an emphasis on understanding how well they predict future performance. Otherwise, many well-intentioned efforts may fall short of their goals of improving patient outcomes after surgery.

REFERENCES

1. Milstein A, Galvin RS, Delbanco SF, et al. Improving the safety of health care: the Leapfrog initiative. *Eff Clin Pract*. 2000;3:313–316.
2. Texas Health Care Information Council. *Indicators of Inpatient Care in Texas Hospitals, 1999–2001*. Austin: Texas Health Care Information Council, 2003.
3. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002;137:511–520.
4. Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *JAMA*. 2000; 283:1159–1166.
5. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
6. Hannan EL. The relation between volume and outcome in health care. *N Engl J Med*. 1999;340:1677–1679.
7. Christian CK, Gustafson ML, Betensky RA, et al. The Leapfrog volume criteria may fall short in identifying high-quality surgical centers. *Ann Surg*. 2003;238:447–455.
8. Rathore SS, Epstein AJ, Volpp KG, et al. Hospital coronary artery bypass graft surgery volume and patient mortality. *Ann Surg*. 2004;239: 110–117.
9. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA*. 2004; 292:847–851.
10. Iezzoni LI, Ash AS, Shwartz M, et al. Predicting in-hospital deaths from coronary artery bypass graft surgery: do different severity measures give different predictions? *Med Care*. 1998;36:28–39.
11. Hannan EL, Racz MJ, Jollis JG, et al. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res*. 1997;31:659–678.
12. Rogowski JA, Horbar JD, Staiger DO, et al. Indirect vs direct hospital quality indicators for very low-birth-weight infants. *JAMA*. 2004;291: 202–209.
13. McClellan M, Staiger DO. Comparing the quality of health care providers. In: Garber A, ed. *Frontiers in Health Policy Research*, vol. 3. Cambridge, MA: MIT Press, 2000:113–136.
14. Department of Health and Human Services. *The International Classification of Diseases*, 9th rev. *Clinical Modification*: *ICD-9-CM*. Washington, DC: Government Printing Office, 1998.
15. Leapfrog Group. Evidence-Based Hospital Referral Fact Sheet. www.leapfroggroup.org/FactSheets.htm Accessed May 25th, 2004.
16. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–383.
17. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol*. 1993;46:1075–1079.
18. Luft HS, Romano PS. Chance, continuity, and change in hospital mortality rates: coronary artery bypass graft patients in California hospitals. *JAMA*. 1993;270:331–337.
19. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127:666–674.
20. Urbach DR, Baxter NN. Does it matter what a hospital is 'high volume' for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ*. 2004;328:737–740.
21. Pronovost PJ, Angus DC, Dorman T, et al. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. *JAMA*. 2002;288:2151–2162.
22. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*. 2002; 288:1987–1993.
23. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003;349:2117–2127.
24. Callahan MA, Christos PJ, Gold HT, et al. Influence of surgical subspecialty training on in-hospital mortality for gastrectomy and colectomy patients. *Ann Surg*. 2003;238:629–636.
25. Auerbach AD, Goldman L. beta-Blockers and reduction of cardiac events in noncardiac surgery: clinical applications. *JAMA*. 2002;287: 1445–1447.
26. Goodney PP, O'Connor GT, Wennberg DE, et al. Do hospitals with low mortality rates in coronary artery bypass also perform well in valve replacement? *Ann Thorac Surg*. 2003;76:1131–1136.