Composite Measures for Profiling Hospitals on Bariatric Surgery Performance

Justin B. Dimick, MD, MPH; Nancy J. Birkmeyer, PhD; Jonathan F. Finks, MD; David A. Share, MD, MPH; Wayne J. English, MD; Arthur M. Carlin, MD; John D. Birkmeyer, MD

IMPORTANCE The optimal approach for profiling hospital performance with bariatric surgery is unclear.

OBJECTIVE To develop a novel composite measure for profiling hospital performance with bariatric surgery.

DESIGN, SETTING, AND PARTICIPANTS Using clinical registry data from the Michigan Bariatric Surgery Collaborative, we studied all patients undergoing bariatric surgery from January 1, 2008, through December 31, 2010. For laparoscopic gastric bypass surgery, we used empirical Bayes techniques to create a composite measure by combining several measures, including serious complications, reoperations, and readmissions; hospital and surgeon volume; and outcomes with other related procedures. Hospitals were ranked for 2008 through 2009 and placed in 1 of 3 groups: 3-star (top 20%), 2-star (middle 60%), and 1-star (bottom 20%). We assessed how well these ratings predicted outcomes in the next year (2010) compared with other widely used measures.

MAIN OUTCOMES AND MEASURES Risk-adjusted serious complications.

RESULTS Composite measures explained a larger proportion of hospital-level variation in serious complication rates with laparoscopic gastric bypass than other measures. For example, the composite measure explained 89% of the variation compared with only 28% for risk-adjusted complication rates alone. Composite measures also appeared better at predicting future performance compared with individual measures. When ranked on the composite measure, 1-star hospitals had 2-fold higher serious complication rates (4.6% vs 2.4%; odds ratio, 2.0; 95% CI, 1.1-3.5) compared with 3-star hospitals. Differences in serious complication rates between 1- and 3-star hospitals were much smaller when hospitals were ranked using serious complications (4.0% vs 2.7%; odds ratio, 1.6; 95% CI, 0.8-2.9) and hospital volume (3.3% vs 3.2%; odds ratio, 0.85; 95% CI, 0.4-1.7).

CONCLUSIONS AND RELEVANCE Composite measures are much better at explaining hospital-level variation in serious complications and predicting future performance than other approaches. In this preliminary study, it appears that such composite measures may be better than existing alternatives for profiling hospital performance with bariatric surgery.

Author Affiliations: The Michigan Bariatric Surgery Collaborative (MBSC), University of Michigan, Ann Arbor (Dimick, N.J. Birkmeyer, Finks, J.D. Birkmeyer); Marquette General Hospital, Marquette, Michigan (English); Henry Ford Health System, Detroit, Michigan (Carlin); Blue Cross Blue Shield of Michigan, Detroit (Share).

Corresponding Author: Justin B. Dimick, MD, MPH, Center for Healthcare Outcomes & Policy, 2800 Plymouth Rd, Bldg 126, Room 137E, Ann Arbor, MI 48109 (jdimick@umich.edu).

Published online October 16, 2013.
There is growing enthusiasm for overhauling the current hospital accreditation program for bariatric surgery. The American College of Surgeons (ACS) and the American Society for Metabolic and Bariatric Surgery (ASMBPS), the 2 leading professional organizations that offer accreditation, currently rely on a volume standard (>125 cases per year) among other structure and process measures. Recent studies, however, have shown that centers receiving accreditation, so-called centers of excellence, do not have better outcomes than other hospitals. As a result, there is mounting pressure to move beyond volume standards toward more direct measures of hospital outcomes.

However, the best approach for profiling hospitals based on outcomes is unclear. Many advocate directly measuring outcomes alone (eg, serious morbidity and reoperation), but these measures may be too “noisy” to reliably reflect performance. Small sample sizes and low event rates conspire to limit the precision of hospital outcome measures. Because of these limitations, there is growing use of composite measures to ascertain hospital performance. Composite measures combine multiple different quality indicators into a single score to increase the reliability of hospital performance assessment. Although composite measures have been applied to other medical and surgical conditions, to our knowledge, their use has not been explored in bariatric surgery.

In this article, we use data from the Michigan Bariatric Surgical Collaborative (MBSC) clinical registry to explore the value of composite measures of bariatric surgery performance. To create these composite measures, we combined multiple measures by weighting each according to its ability to predict serious morbidity. Because bariatric surgery accreditation is used for selective referral (eg, with the Centers for Medicare & Medicaid Services [CMS] national coverage decision), we evaluated this composite measure by its ability to predict future performance compared with other widely used measures. The ability to predict future performance is an essential criterion for quality measures used for selective referral since patients and payers make decisions about where to have surgery based on historical outcomes data.

Methods

Data Source and Study Population

This study is based on data from the MBSC, a payer-funded quality improvement program that administers a prospective, externally audited outcomes registry of patients undergoing bariatric surgery in Michigan. The MBSC is a consortium of 29 Michigan hospitals and 75 surgeons performing bariatric surgery and has been described in detail elsewhere. Participation in the MBSC is voluntary, and any hospital that performs at least 25 bariatric procedures per year is eligible to volunteer. The MBSC currently enrolls approximately 6000 patients annually into its clinical registry. Procedures meeting this definition include open and laparoscopic gastric bypass, adjustable gastric banding, sleeve gastrectomy, and bilipancreatic diversion with duodenal switch. Participating hospitals submit data on all patients undergoing primary and revisional bariatric surgery.

For the MBSC clinical registry, data on patient characteristics, procedure type, processes of care, and postoperative outcomes are obtained by medical record abstraction at the end of the perioperative period (in hospital and up to 30 days after surgery). The medical records are reviewed by centrally trained data abstractors using a standardized and validated data collection instrument. Each hospital within the MBSC is audited annually by nurses from the coordinating center to verify the accuracy and completeness of its clinical registry data. For this study, we identified all patients undergoing a primary (non-revisional) bariatric surgical procedure from January 1, 2008, through December 31, 2010. We excluded patients undergoing a revisional procedure because of the heterogeneity of risk in this patient population.

Outcomes Assessment

The MBSC registry includes clinical data on 13 different types of perioperative complications. Complications were determined by documentation of the specific complication in the medical record, including confirmatory radiographic imaging reports when available, as well as the treatment for the complication. In MBSC, complications are categorized according to severity as non–life-threatening (grade 1), potentially life-threatening (grade 2), or life-threatening complications associated with permanent residual disability or death (grade 3).

For the purposes of this study, we included all serious complications (grade 2 or higher) in our primary outcome variable. Grade 2 complications included abdominal abscess (requiring percutaneous drainage or reoperation), bowel obstruction (requiring reoperation), leak (requiring percutaneous drainage or reoperation), bleeding (requiring blood transfusion of >4 U, endoscopy, reoperation, or splenectomy), wound infection or dehiscence (requiring reoperation), respiratory failure (requiring 2–7 days of mechanical ventilation), renal failure (requiring in-hospital dialysis), venous thromboembolism (deep venous thrombosis or pulmonary embolism), and band-related problems requiring reoperation (port site infection, gastric perforation, band slippage, and outlet obstruction). Grade 3 complications included myocardial infarction or cardiac arrest, renal failure requiring long-term dialysis, respiratory failure (requiring >7 days of mechanical ventilation or tracheostomy), and death.

Hospital Quality Rankings

Our goal was to compare the composite measure with several existing approaches for assessing hospital performance with bariatric surgery. Herein we describe the methods used to rank hospitals on the following measures: (1) hospital volume, (2) risk-adjusted complication rates, (3) risk- and reliability-adjusted complication rates, and (4) composite measures.

Hospital Volume

We calculated hospital volume by assessing the mean number of cases per year from 2009 through 2010. For hospitals that contributed data for less than 2 years, we estimated their predicted annual volume by multiplying their monthly volume by the total number of months they submitted data. All included hospitals had data for more than 12 months.
Risk-Adjusted Complication Rates

We used standard techniques for calculating risk-adjusted complication rates for each hospital. Data on patient characteristics included demographics (age, sex, and payer type), height, weight, mobility limitations (requiring ambulation aids, non-ambulatory, or bed-bound), smoking status, and comorbid conditions. The height and weight were used to determine body mass index, calculated as weight in kilograms divided by height in meters squared. The definitions for most comorbidities included documentation of the condition and its treatment in the medical record. Comorbid conditions included pulmonary disease (asthma, obstructive or restrictive disorders, home oxygen use, or Pickwickian syndrome), cardiovascular disease (coronary artery disease, dysrhythmia, peripheral vascular disease, stroke, hypertension, or hyperlipidemia), sleep apnea, psychological disorders, prior venous thromboembolism, type 2 diabetes mellitus, chronic renal failure (requiring dialysis or transplant), liver disease (nonalcoholic fatty liver disease, cirrhosis, or liver transplantation), urinary incontinence, gastroesophageal reflux disease, peptic ulcer disease, cholelithiasis, previous ventral hernia repair, and musculoskeletal disorders.

Logistic regression models were used to create models that included all significant patient-level covariates. The predicted probabilities of each patient were estimated from this model and then summed for each hospital to calculate the “expected” number of deaths. The observed number of deaths was then divided by the expected number to yield an observed to expected ratio. This ratio was then multiplied by the overall mean to yield a risk-adjusted morbidity rate for each hospital.

Risk- and Reliability-Adjusted Complication Rates

Reliability adjustment is an increasingly used technique to adjust hospital-specific outcomes for statistical noise. This is based on the standard shrinkage estimator approach that places more weight on a hospital’s complication rate when it is measured reliably but shrinks back toward the mean complication rate when a hospital’s complication rate is measured with error (eg, for hospitals with small numbers of patients undergoing the procedure). For this study, we performed reliability adjustment by generating empirical Bayes estimates of hospital-specific risk-adjusted complication rates for each hospital. To create these estimates, we first used a hierarchical model in which the patient was the first level and the hospital was the second level. The dependent variable was complications at the patient level, and the same risk adjustment variables described earlier were included at the first level as independent variables. The second level included only a hospital-level random effect. The random effect in log(odds) was then predicted using empirical Bayes techniques. This was added back to the mean log(odds) of complications in the overall population, and an inverse logit was performed to estimate the risk and reliability-adjusted complication rate for each hospital.

Composite Measure

We developed a composite measure that incorporates information from multiple quality indicators to optimally predict “true” risk-adjusted complications for laparoscopic gastric bypass. In creating these measures, we considered several individual quality measures, including hospital volume and several risk-adjusted outcomes (mortality, complications, reoperation, readmission, and length of stay). We also considered risk-adjusted outcomes not only for the index operation but also for other related procedures (eg, laparoscopic gastric banding and sleeve gastrectomy). Of note, all input measures were risk-adjusted using clinical registry data, as described earlier in the section on risk-adjusted complication rates.

Our composite measure is a generalization of the reliability adjustment described previously. While the simple shrinkage estimator is a weighted mean of a single measure of interest, our composite measure is a weighted mean of all available quality indicators. The weight on each quality indicator is determined for each hospital to minimize the expected mean squared prediction error, using an empirical Bayes method. Although the statistical methods used to create these measures are described in detail elsewhere, we will provide a brief conceptual overview. The first step in creating the composite measure was to determine the extent to which each quality indicator predicts risk-adjusted complication rates for the index operation. To evaluate the importance of each potential input, we first estimated the proportion of systematic (ie, nonrandom) variation explained by each quality indicator (Table 1).

We included any quality indicator in the composite measure that explained more than 10% of hospital variation in risk-adjusted complications from 2008 through 2009.

Next, we calculated weights for each quality indicator. The weight placed on each quality indicator in our composite measure was based on 2 factors. The first is the hospital-level correlation of each quality indicator with the complication rate for the index operation. The strength of these correlations indicates the extent to which other quality indicators can be used to help predict complications for the index operation. The second factor affecting the weight placed on each quality indicator is the reliability with which each indicator is measured. Reliability ranges from 0 (no reliability) to 1 (perfect reliability). The reliability of each quality indicator refers to the proportion of the overall variance attributable to true hospital-level variation in performance, as opposed to estimation error (noise). For example, in smaller hospitals, less weight is placed on complication rates because they are less reliably estimated.
Statistical Analysis
We determined the value of different quality indicators by calculating how well hospital performance rankings from 2008 through 2009, based on each measurement approach, predicted risk-adjusted serious complications in the next year (2010). For each operation, hospitals were ranked based on each quality measure (data from 2008-2009) and were divided into 3 even-sized groups (1-, 2-, and 3-star). The “worst” hospitals (bottom 20%) received a 1-star rating, the middle of the distribution (middle 60%) received a 2-star rating, and the “best” hospitals (top 20%) received a 3-star rating. We then assessed the ability of our composite measure to predict future performance compared with standard techniques for ranking hospitals, including hospital volume, risk-adjusted complication rates, risk- and reliability-adjusted complication rates, and our composite measure. For these analyses, we evaluated the discrimination in future risk-adjusted complications, comparing the 1-star hospitals with the 3-star hospitals for each measure. These analyses were conducted using patient-level logistic regression models in the 2010 data. The dependent variable was 1 or more serious complications, and the independent variables were patient characteristics used for risk adjustment. Each quality indicator from the 2008 through 2009 data was then added as a dummy variable (1-, 2-, and 3-star), and we present the odds ratio and 95% CI representing 3- vs 1-star hospitals.

We also assessed the ability of the composite measure and other quality indicators (assessed in 2008-2009) to explain future (2010) hospital-level variation in risk-adjusted serious complications. To avoid problems with noise variation in the subsequent period, we determined the proportion of systemic hospital-level variation explained. We generated hierarchical models with 1 or more complications as the dependent variable (2009) and used them to estimate the hospital-level variance. We first used an “empty model” that contained only patient variables for risk adjustment. Then, we entered each historical quality indicator (assessed in 2008-2009) into the model. Next, we calculated the degree to which the historical quality indicators reduced the hospital-level variance, an approach described in our prior work. All statistical analyses were conducted using STATA, version 11.0 (StataCorp).

Results
Inputs to the Composite Measure
Several individual quality indicators explained hospital-level variation in serious complications with bariatric surgery, varying from hospital volume, which explained 11%, to rates of prolonged length of stay, which explained 65% (Table 1).

The weights applied to each quality indicator in the composite measure are shown in Table 2. The largest amount of weight (39%) is applied to hospital structural characteristics and the overall mean (ie, the target to which observed performance is anchored and “shrunk” back toward). Complications received the next highest weight (25%), followed by readmission (18%), prolonged length of stay (10%), emergency department revisit (6%), reoperation (6%), and complications with other procedures, such as sleeve gastrectomy (2%) and laparoscopic gastric banding (2%) (Table 2).

Patient Characteristics
When hospitals were grouped according to the composite measures, there were no substantial differences in patient illness levels across the groups (Table 3). Although several individual characteristics differed significantly between groups, the expected mortality—a function of all patient characteristics combined—was the same (3.3%) across centers (Table 3).

Evaluation of the Composite Measure
The composite measure predicted larger differences in future performance compared with the other quality indicators (Figure). The spread from 3- (top 20%) to 1-star (bottom 20%) hospitals in future risk-adjusted complication rates for each measure was as follows: hospital volume, 3.3% to 3.2%; risk-adjusted complications, 2.7% to 4.0%; risk- and reliability-adjusted complications, 2.4% to 4.1%; and the composite measure, 2.4% to 4.6%. In the logistic regression model, which accounts for all patient characteristics, the composite measure was the only quality indicator that predicted statistically significant differences between 1- and 3-star hospitals (Table 4).

The composite measure also explained a higher proportion of hospital-level variation than the other quality measures (Table 4). This analysis confirmed the same trend shown earlier for discrimination among 1- and 3-star hospitals. Hospital volume explained no variation (0%), and risk-adjusted complications (28%) and risk- and reliability-adjusted complications (47%) explained a larger fraction. However, the composite measure explained the largest proportion of hospital-level variation in serious complications (89%).

Discussion
In this article, we demonstrate that a composite measure of bariatric surgery performance is superior to existing quality indicators at identifying the hospitals with the best outcomes. The composite measure described in this study is created by combining multiple different outcomes (eg, complications, reoperation, and

Table 2. Weights Placed on Each Quality Indicator in the Composite Measure of Performance for Laparoscopic Gastric Bypass

<table>
<thead>
<tr>
<th>Hospital Quality Indicator</th>
<th>Weight Afforded in the Composite Measure, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected complication rate given hospital structural characteristics and overall mean complication rate</td>
<td>39</td>
</tr>
<tr>
<td>Complication rate with laparoscopic gastric bypass</td>
<td>25</td>
</tr>
<tr>
<td>Readmission with laparoscopic gastric bypass</td>
<td>18</td>
</tr>
<tr>
<td>Prolonged length of stay with laparoscopic gastric bypass</td>
<td>10</td>
</tr>
<tr>
<td>Emergency department revisit rate with laparoscopic gastric bypass</td>
<td>6</td>
</tr>
<tr>
<td>Reoperation with laparoscopic gastric bypass</td>
<td>6</td>
</tr>
<tr>
<td>Complication rates with sleeve gastrectomy</td>
<td>2</td>
</tr>
<tr>
<td>Complication rates with laparoscopic gastric banding</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3. Patient Characteristics for the 1-, 2-, and 3-Star Hospitals According to Hospital Rankings Created Using the Composite Measure*

<table>
<thead>
<tr>
<th>Patient Characteristic</th>
<th>3-Star (n = 1033)</th>
<th>2-Star (n = 968)</th>
<th>1-Star (n = 941)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected complication rate, mean</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Age &gt;50 y</td>
<td>37</td>
<td>42</td>
<td>36</td>
</tr>
<tr>
<td>Male sex</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Body mass index, meanb</td>
<td>47.7</td>
<td>49.4</td>
<td>48.6</td>
</tr>
<tr>
<td>History of thromboembolism</td>
<td>3.2</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Smoking history</td>
<td>38</td>
<td>42</td>
<td>46</td>
</tr>
<tr>
<td>Pulmonary disease</td>
<td>34</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td>6.4</td>
<td>5.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Mobility limitations</td>
<td>9.0</td>
<td>5.5</td>
<td>8.8</td>
</tr>
<tr>
<td>Sleep apnea</td>
<td>50</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>Hypertension</td>
<td>65</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>Type 2 diabetes mellitus</td>
<td>41</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>57</td>
<td>59</td>
<td>55</td>
</tr>
<tr>
<td>Liver disease</td>
<td>3.0</td>
<td>5.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Gastroesophageal reflux disease</td>
<td>56</td>
<td>48</td>
<td>57</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>5.3</td>
<td>1.9</td>
<td>4.3</td>
</tr>
<tr>
<td>Musculoskeletal disorder</td>
<td>80</td>
<td>77</td>
<td>82</td>
</tr>
<tr>
<td>Prior hernia repair</td>
<td>2.8</td>
<td>2.2</td>
<td>1.6</td>
</tr>
</tbody>
</table>

* Values are presented as percentages unless otherwise indicated.

b Calculated as weight in kilograms divided by height in meters squared.

Figure. Future Risk-Adjusted Mortality Rates (2010) for 3-Star (Top 20%), 2-Star (Middle 60%), and 1-Star (Bottom 20%) Hospitals


<table>
<thead>
<tr>
<th>Hospital Ranking (2008-2009)</th>
<th>Risk-Adjusted Complications in 2010</th>
<th>Odds Ratio (95% CI)</th>
<th>Proportion of Hospital-Level Variation Explained, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- vs 3-Star Hospitals</td>
<td>1- vs 3-Star Hospitals</td>
<td></td>
</tr>
<tr>
<td>Hospital volume</td>
<td>0.93 (0.47-1.85)</td>
<td>0.85 (0.43-1.68)</td>
<td>0</td>
</tr>
<tr>
<td>Risk-adjusted complication rates</td>
<td>1.27 (0.66-2.44)</td>
<td>1.56 (0.84-2.91)</td>
<td>28</td>
</tr>
<tr>
<td>Risk- and reliability-adjusted complication rates</td>
<td>1.37 (0.71-2.61)</td>
<td>1.65 (0.88-3.10)</td>
<td>47</td>
</tr>
<tr>
<td>Composite measure</td>
<td>1.26 (0.7-2.25)</td>
<td>1.99 (1.14-3.47)</td>
<td>89</td>
</tr>
</tbody>
</table>
Hospital Performance With Bariatric Surgery

Original Investigation Research


ARTICLE INFORMATION

Accepted for Publication: March 7, 2013.

Published Online: October 16, 2013.

Author Contributions: Study concept and design: Dimick, N. J. Birkmeyer, J. D. Birkmeyer.

Acquisition of data: Dimick, N. J. Birkmeyer, J. D. Birkmeyer.

Analysis and interpretation of data: All authors.

Drafting of the manuscript: Dimick, N. J. Birkmeyer, J. D. Birkmeyer.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Dimick, N. J. Birkmeyer, J. D. Birkmeyer.

Obtained funding: Dimick, Share.

Administrative, technical, and material support: Dimick, Finks.

Study supervision: Dimick.

Copyright 2014 American Medical Association. All rights reserved.

prolonged length of stay), structural variables, and outcomes with other related bariatric surgery procedures. The composite measure was better at predicting future performance and explained a higher proportion of hospital-level variation than the most widely used quality indicators for bariatric surgery.

Composite measures are becoming more widely used to profile hospital performance in other areas of surgery. For example, the Society of Thoracic Surgeons, which maintains a clinical registry that captures nearly all cardiac surgery programs in the United States, uses a composite measure for profiling hospitals.15,16,17 The measure is a combination of processes of care and multiple outcomes (eg, death and complications). It has 1 key conceptual difference compared with our approach—it is designed to reflect “global” quality across all domains of performance. Because of this objective, each domain (death, complications, and processes of care) is afforded equal weight. In contrast, our goal was to combine the quality signal from multiple measures to best predict a single criterion-standard outcome, risk-adjusted serious complications.8

In addition to the literature supporting the use of composite measures for assessing hospital performance, there is also a growing use of so-called reliability adjustment. These approaches address measurement problems in small hospitals by shrinking an observed outcome back toward the mean in the population. While there is consensus that this approach is better than using noisy outcome measures, there is a great deal of debate about which value should be the target for shrinkage.8,17 For example, with the CMS Hospital Compare measures for acute myocardial infarction, heart failure, and pneumonia, the risk-adjusted mortality and readmission rates are shrunk toward the mean rate for all hospitals. Many have challenged this approach because it assumes that small hospitals have average performance.17 For any procedure or condition with a strong volume-outcome relationship, however, this assumption clearly does not hold.

The method used in this article provides an alternative approach. Rather than making the assumption that all hospitals have average performance, we use a flexible approach that incorporates information on hospital and surgeon characteristics, including hospital volume. With such a strategy, the relationship between hospital characteristics and the outcome of interest is explicitly modeled and incorporated. For example, if hospital volume is an important predictor of outcomes, these methods would shrink the hospital outcome rate toward the expected outcome for the hospital volume group. Small hospitals with very low case loads and little “signal” would therefore have an outcome rate close to the rate at low-volume hospitals. In the present study, hospital volume was not a strong predictor of complications and therefore did not receive much weight. However, in other studies of surgical procedures (eg, esophageal and pancreatic cancer resections), hospital volume was an extremely important input to the composite measure.8,9

The present study has several limitations. First, our registry is restricted to hospitals in Michigan, which may not be representative of the nation as a whole. Specifically, the relationship between hospital volume and outcomes may be stronger in states without a regional quality improvement program. With a national sample of hospitals, it is possible that our results would be different. However, this approach may change the relative weight placed on input measures, but it is unlikely to affect our main findings—that a composite measure that combines signal across measures is superior to individual measures alone. Second, the present study focuses on short-term outcomes, such as perioperative safety. Quality in bariatric surgery is much more broad than short-term safety and should include measures of longer-term effectiveness, such as weight loss, comorbid disease resolution, and patient satisfaction. Unfortunately, however, long-term follow-up data are not widely available. Where they do exist, they are often incomplete or inaccurate. Future accreditation efforts should emphasize the complete collection of these long-term data. Finally, our study is limited because it includes a relatively small sample of hospitals. This sample size restriction prevents us from performing a bootstrapping (or resampling) analysis to directly compare the composite measure with the other measures. As such, our study should be viewed as preliminary and exploratory in nature, and it needs to be replicated in a national cohort of bariatric surgery patients.

The findings of this study have important policy implications. Bariatric surgery is one of only a few surgical procedures for which accreditation is currently linked to insurance coverage. In 2006, the CMS issued a national coverage decision for bariatric surgery that limited payment for surgery performed in hospitals accredited by the ACS and the ASMBS. Many private payers have since linked coverage or created tiered networks that steer patients toward ACS- and ASMBS-accredited centers. However, recent evidence from clinical registry and administrative data has shown that accredited centers do not have better performance compared with nonaccredited centers. Better measures of hospital quality are needed to ensure that selective referral efforts of the CMS and private payers are having the intended effect—steering patients to safer hospitals. More reliable measures are also needed for benchmarking and quality improvement in bariatric surgery. The ACS and ASMBS are currently moving away from the “center of excellence” model and developing a national outcomes feedback program. Reliable outcome measures are needed to give hospitals and surgeons a true sense of where they stand compared with their peers. This study provides preliminary data that empirically weighted composite outcomes measures may be better than existing alternatives for selective referral and outcomes feedback programs.
Hospital Quality

Does Past Performance Predict Future Performance?

Laurent G. Glance, MD; Andrew W. Dick, PhD; Turner M. Osler, MD, MS; Arthur L. Kellermann, MD, MPH

More than 10 years after the Institute of Medicine told Americans that medical errors are the third leading cause of death in this country,1 the federal government, third-party payers, and patients are no longer trying to do. Instead, we rely on complex risk adjustment models that predict individual patient outcomes as a function of baseline risk factors. We benchmark quality by estimating whether the observed outcomes of a patient cohort treated by a particular hospital (or physician) are better or worse than the “average” hospital.4

Before risk adjustment models are used for quality measurement, they are validated by determining how well their predictions agree with observed outcomes. In their article, Dimick and colleagues5 examine model performance using a different approach. Rather than follow the conventional path, they compared how well different quality metrics generated from the Michigan Bariatric Surgery Collaborative.6

1 2010;304(4):435-442.