# Title Page

Project Title: Composite Measures of Surgical Performance

Principal Investigator: Justin B. Dimick, MD, MPH

Team Members: John Birkmeyer, MD (Mentor)

Organization: The Regents of the University of Michigan

Inclusive Dates of the Project: 09/30/2008 – 06/30/2013

Federal Project Officer: Kay Anderson

Acknowledgment of Agency Support: Agency for Healthcare Research and Quality (AHRQ)

Grant Number: K08HS017765

## Structured Abstract

**Purpose:** To develop and validate empirically weighted composite measures of surgical morbidity.

**Scope:** Each year in the United States, more than 20 million patients undergo inpatient surgery. Approximately 5-10% of patients will experience a complication, leading to 1-2 million patients with major surgical morbidity.

**Methods:** Using data from the ACS NSQ Improvement Program, we studied patients undergoing four procedures (2008 to 2009): colectomy, ventral hernia repair, abdominal aortic aneurysm repair, and lower-extremity bypass surgery. We created a composite measure by combining quality indicators from several distinct domains of quality: morbidity, reoperation, length of stay, and morbidity with other, potentially related procedures. To validate this approach, we assessed how well measures from 2008 could predict morbidity in 2009.

**Results:** For all four operations, the composite measures explained a higher proportion of hospital-level variation in morbidity than the standard approach: ventral hernia repair (58% for the composite vs. 8% for the standard approach), colon resection (33% vs. 14%), abdominal aortic aneurysm repair (51% vs. 38%), and lower-extremity bypass surgery (32% vs. 3%). When evaluating the ability to discriminate future performance, the composite performed best for all procedures. For example, with ventral hernia repair, the bottom 20% of hospitals based on the composite had nearly 3-fold higher (OR, 2.65; 95% CI, 1.83 to 3.85) morbidity rates compared to the top 20% of hospitals. However, when using the standard approach, there was only a 1.3-fold difference (OR, 1.30; 95% CI, 0.87 to 1.96).


**Key Words:** Surgery, Quality, Composite, Measures, Outcomes, Process

**Purpose**.

## Aim 1. To develop and validate composite measures of surgical morbidity.

The NSQIP collects information on postoperative morbidity based on standardized definitions. Using this data, we will develop empirically derived composite measures of performance for each operation. Model inputs will include measures from different quality domains, including mortality rates (when relevant) and morbidity rates with the operation of interest and morbidity and mortality rates with related operations. Model outputs will be hospital- and procedure-specific estimates of "true" morbidity that optimally filter out statistical noise. We will validate these composite measures by determining the extent to which they explain hospital-level variation and forecast subsequent outcomes.

## Aim 2. To establish the value of incorporating process of care variables into composite measures.

The NSQIP has begun routinely collecting data on the process measures set forth in the Surgical Care Improvement Program (SCIP). These processes pertain to details of perioperative care (e.g., prophylaxis for infectious, thrombotic, respiratory, and cardiac complications). Using this data, we will first establish the relationship between each process of care and hospital level variations in morbidity. We will then determine whether addition of the SCIP process measures to the composite morbidity measure improves its ability to explain hospital-level variation and forecast subsequent outcomes.

**Scope**

## Surgical morbidity and mortality as a public health problem

Risk-adjusted morbidity is widely used as an indicator of hospital performance with surgery; it is also used by the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP). Although clinically intuitive for surgeons, simple rates of risk-adjusted morbidity may not reliably reflect hospital performance with surgery. Because of low event rates or low hospital caseloads, hospital morbidity rates may be too imprecise (i.e., "noisy) to correctly identify high- and low-performing hospitals.

Composite measures may be a more effective approach for capturing a hospital's quality with surgical care. Compared to rates of risk-adjusted morbidity, composite measures are more effective at addressing problems with statistical "noise." By combining multiple quality indicators for a single operation (e.g., morbidity, length of stay, reoperation), this approach strengthens the quality signal and improves reliability. Moreover, composite measures can further improve precision by adding quality information from other, related procedures. Prior studies demonstrate the superiority of these techniques for profiling hospitals on mortality, but it is unclear whether this approach will also be useful for risk-adjusted morbidity.

In this context, we sought to evaluate whether composite measures could be used to improve the reliability of risk-adjusted morbidity. Using data from the ACS-NSQIP, we developed and evaluated composite measure for several common, high-risk procedures. Each measure was developed by empirically weighting several input measures, including quality indicators for the index operation and other potentially related operations. We then assessed the ability of these measures to explain systematic variation in hospital-level morbidity and predict future risk-adjusted morbidity compared to simple rates of risk-adjusted hospital morbidity.

## Methods

### Data Source and Study Population

We used data from the 2008 and 2009 American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP). The ACS-NSQIP is a prospective, multi-institutional clinical registry created to feed back risk-adjusted outcomes to hospitals for quality improvement purposes. We included all participating centers with data for both 2008 and 2009. Over 130 pre- and post-operative variables are recorded, including patient demographics, preoperative risk factors, patient laboratory values, intraoperative variables, and postoperative 30-day morbidity and mortality. The data collection process relies on a sampling strategy aimed at collecting a diverse set of operations. Trained surgical clinical nurse reviewers record the data using standardized definitions. The accuracy of the data is ensured through intensive training mechanisms for the surgical clinical nurse reviewers and inter-rater reliability audits of participating sites. For this study, we used appropriate Current Procedure Terminology (CPT) codes to identify all patients undergoing one of four common, high-risk procedures: (ventral hernia repair, abdominal aortic aneurysm repair, colectomy, and pancreatectomy) at participating hospitals.

### Hospital Morbidity Rates

We used standard ACS-NSQIP techniques for calculating risk-adjusted morbidity rates for each hospital. For the purposes of this study, we limited our assessment to serious morbidity, which included organ space infection, wound dehiscence, pneumonia, unplanned reintubation, postoperative bleeding, stroke, acute myocardial infarction, acute renal failure, sepsis, septic shock, deep venous thrombosis, and pulmonary embolism. Data on complications were ascertained by trained nurse clinicians according to standardized definitions. Hospital morbidity rates were risk-adjusted using detailed data on patient characteristics. For this risk-adjustment, stepwise logistic regression is 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 “O/E ratio.” This ratio was then multiplied by the overall average to yield a risk-adjusted morbidity rate for each hospital.

**Composite Morbidity Measure**

We developed a composite measure that incorporates information from multiple quality indicators to optimally predict “true” risk-adjusted morbidity for each operation. In creating these measures, we considered several individual quality measures, including morbidity rates, reoperation, and length of stay. For each operation, we considered morbidity not only for the index operation but also for other, related procedures (e.g., colectomy morbidity rates were tested as inputs to the composite measure for other general surgery procedures).

Our composite measure is a generalization of the standard shrinkage estimator that places more weight on a hospital’s own morbidity rate when it is measured reliably, but shrinks back toward the average morbidity when a hospital’s own morbidity is measured with error (e.g., for hospitals with small numbers of patients undergoing the procedure). The simple shrinkage estimator is a weighted average of a single measure of interest and its mean; our composite measure is a weighted average of all available quality indicators – the morbidity rates for all procedures that are thought to be potentially related. The weight on each quality indicator is determined for each hospital to minimize the expected mean squared prediction error, using an empirical Bayes methodology.

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 individual quality indicator predicts risk-adjusted morbidity for the index operation. To evaluate the importance of each potential input, we first estimated the proportion of systematic (i.e., nonrandom) variation in risk-adjusted morbidity explained by each individual quality indicator (Table 1). We included any quality indicator in the composite measure that explained more than 10% of hospital variation in risk-adjusted morbidity during 2008.

Next, we calculated weights for each quality indicator. The weight placed on each quality indicator in our composite measure was based on two factors. The first is the hospital-level correlation of each quality indicator with the morbidity rate for the index operation. The strength of these correlations indicates the extent to which other quality indicators can be used to help predict morbidity 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 refers to the proportion of the overall variance that is attributable to true hospital-level variation in performance, as opposed to estimation error ("noise"). For example, in smaller hospitals, less weight is placed on mortality and morbidity rates because they are less reliably estimated. We assume that structural characteristics of each hospital (such as hospital volume) are not estimated with error and, therefore, have reliability equal to 1.

**Analysis**

We determined the value of our composite measure by determining how well it predicted risk-adjusted morbidity in the next year (2009). For each operation, hospitals were ranked based on the composite measure (data from 2008) and assigned one of three rankings (1-star, 2-star, and 3-star). The "worst" hospitals (bottom 20%) received a 1-star rating, the middle of the distribution (60%) received a 2-star rating, and the "best" hospitals (top 20%) received a 3-star rating. Many hospital rating systems determine tiers of performance by designating high and low outliers by testing for statistically significant differences from the average. Because we used empirical Bayes methods, which adjust each hospital's composite for imprecision (i.e., hospital rankings are a valid indicator of relative performance), we used percentile cutoffs. We then calculated the risk-adjusted mortality rates for 1-star, 2-star, and 3-star hospitals during the subsequent two years (data from years 2007-08). We next assessed the ability of our composite measure to predict future performance compared to standard techniques for ranking hospitals on risk-adjusted morbidity. For these analyses, we evaluated the discrimination in future, risk-adjusted morbidity, comparing the 1-star hospitals (bottom 20%) to the 3-star hospitals (top 20%) for each of the measures.

We also assessed the ability of the composite measure and standard risk-adjusted morbidity (assessed in 2008) to explain future (2009), hospital-level variation in risk-adjusted morbidity. To avoid problems with "noise variation" in the subsequent time period, we determined the proportion of systemic hospital-level variation explained. We generated hierarchical models with morbidity 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. We then entered each historical quality measure (assessed in 2008) into the model. We then calculated the degree to which the historical quality measures reduced the hospital-level variance, an approach described in our prior work. All statistical analyses were conducted using STATA 10.0 (College Station, Texas).

## Results

**Inputs to the composite measure.** For each of the four procedures, several individual measures explained a significant proportion of hospital-level variation in risk-adjusted morbidity (Table 1). The amount of hospital-level variation explained by each procedure's own morbidity rate varied, ranging from 60% with colon resection to only 17% for abdominal aortic aneurysm repair (Table 1). Morbidity with other, related procedures was important in explaining hospital-level variation for all four procedures (Table 1). For example, morbidity with colectomy and pancreatectomy explain 26% and 36% of the hospital-level variation in risk-adjusted morbidity with ventral hernia repair, respectively.

Hospital length of stay with the index procedure also explained a large proportion of hospital-level variation in morbidity, varying from 29% with abdominal aortic aneurysm repair to 11% with lower extremity bypass surgery (Table 1). Similarly, the hospital reoperation rate explained up to 21% of hospital-level morbidity with colon resection, but only 5% for abdominal aortic aneurysm repair.

**Ability of the composite to explain hospital-level variation.** The composite measures explained a high proportion of systematic hospital-level variation in subsequent risk-adjusted morbidity (Table 3). For each operation, the composite measure explained a much higher proportion of variation than did the standard approach to measuring morbidity: ventral hernia repair (58% vs. 8%), colon resection (33% vs. 14%), abdominal aortic aneurysm repair (51% vs. 38%), and lower-extremity bypass surgery (32% vs. 3%) (Table 3).

**Ability of the composite to predict future performance.** The composite score, created by combining these individual measures, performed well at predicting future hospital performance (Table 3, Figure 1). For all four procedures, the composite measure based on 2008 data was better at discriminating future performance in 2009 than the standard approach was to measure risk-adjusted morbidity (Table 3, Figure 1). For example, with ventral hernia repair, historical risk-adjusted morbidity predicted a smaller difference between the best (bottom 20%) and worst (top 20%) hospitals (OR, 1.30; 95% CI, 0.87 to 1.96) when compared to the composite measure (OR, 2.65; 95% CI, 1.83 to 3.85) (Table 3). These differences in mortality could not be explained by differences in patient severity of illness, as the differences in patient characteristics, shown in Table 2, were adjusted for in all comparisons.

**Table 1.** Components of the composite measure are shown, along with the proportion of nonrandom hospital-level morbidity explained by each.

| Procedure | Individual quality measures | Proportion of hospital-level variation explained |
|---|---|---|
| **Ventral hernia repair** | **Index operation** | |
| | Morbidity rate | 40% |
| | Length of stay | 15% |
| | Reoperation rate | 12% |
| | **Other operations** | |
| | Morbidity with colectomy | 26% |
| | Morbidity with esophagectomy | 12% |
| | Morbidity with liver resection | 13% |
| | Morbidity with pancreatectomy | 36% |
| **Colon resection** | **Index operation** | |
| | Morbidity rate | 60% |
| | Length of stay | 16% |
| | Reoperation rate | 21% |
| | **Other operations** | |
| | Morbidity with appendectomy | 12% |
| | Morbidity with cholecystectomy | 13% |
| | Morbidity with liver resection | 15% |
| | Morbidity with pancreatectomy | 10% |
| | Morbidity with proctocolectomy | 11% |
| | Morbidity with ventral hernia repair | 15% |
| **Abdominal aortic aneurysm repair** | **Index operation** | |
| | Morbidity rate | 17% |
| | Length of stay | 29% |
| | Reoperation | 5% |
| | **Other operations** | |
| | Morbidity with ventral hernia repair | 11% |
| **Lower extremity bypass surgery** | **Index operation** | |
| | Morbidity rate | 41% |
| | Length of stay | 11% |
| | Reoperation rate | 13% |
| | **Other operations** | |
| | Morbidity with gastric bypass | 19% |

**Table 2.** Patient characteristics for the best, middle, and worst hospitals in 2008.

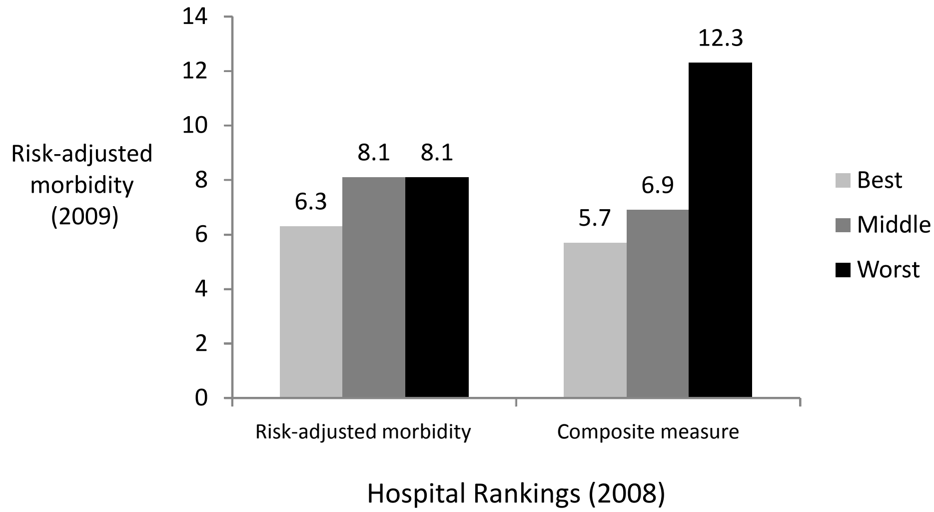
| | Hospitals Ranked on 2008 Composite Measures | | |
|---|---|---|---|
| **Procedure** | Worst 20% of hospitals | Middle 60% of hospitals | Best 20% of hospitals |
| **Ventral hernia repair** | | | |
| Age, mean | 60 | 58 | 56 |
| Non-White race, % | 23 | 23 | 23 |
| Male, % | 39 | 37 | 42 |
| Emergent surgery, % | 8.0 | 12.6 | 11.6 |
| Expected morbidity rates, % | 6.1 | 7.1 | 6.7 |
| **Colon resection** | | | |
| Age, mean | 63 | 63 | 60 |
| Non-White race, % | 22 | 21 | 29 |
| Male, % | 50 | 53 | 53 |
| Emergent surgery, % | 14.7 | 18.0 | 17.6 |
| Expected morbidity rates, % | 15.2 | 16.7 | 16.3 |
| **Elective abdominal aortic aneurysm repair** | | | |
| Age, mean | 73 | 73 | 73 |
| Non-White race, % | 13 | 15 | 17 |
| Male, % | 81 | 78 | 80 |
| Emergent surgery, % | 8.2 | 12.3 | 13.4 |
| Expected morbidity rates, % | 12.1 | 13.8 | 15.2 |
| **Lower extremity bypass surgery** | | | |
| Age, mean | 68 | 66 | 66 |
| Non-White race, % | 21 | 22 | 38 |
| Male, % | 63 | 63 | 61 |
| Emergent surgery, % | 6.4 | 6.0 | 7.4 |
| Expected morbidity rates, % | 10.7 | 10.7 | 10.6 |

**Table 3.** Relative ability of the composite measure and standard risk-adjusted morbidity from 2008 to forecast risk-adjusted morbidity in 2009.

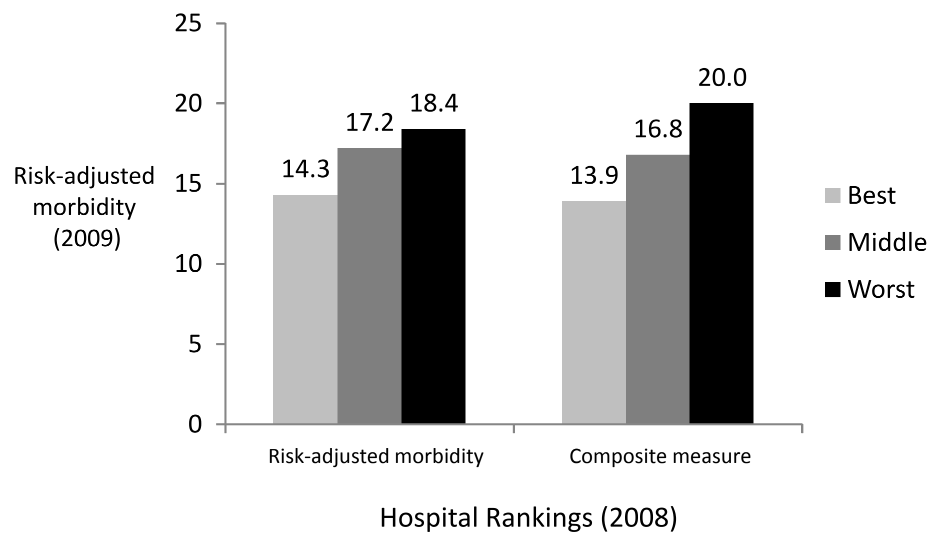
| 2008 Hospital Rankings | 2009 Risk-Adjusted Morbidity | |
|---|---|---|
| | **Odds Ratio, "best" vs. "worst" hospitals (95% CI)** | **% Hospital-Level Variation Explained** |
| **Ventral hernia repair** | | |
| Composite measure | 2.65 (1.83 to 3.85) | 58% |
| Risk-adjusted morbidity | 1.30 (0.87 to 1.96) | 8% |
| **Colon resection** | | |
| Composite measure | 1.70 (1.41 to 2.04) | 33% |
| Risk-adjusted morbidity | 1.47 (1.21 to 1.78) | 14% |
| **Abdominal aortic aneurysm repair** | | |
| Composite measure | 1.72 (1.20 to 2.45) | 51% |
| Risk-adjusted morbidity | 1.35 (0.95 to 1.92) | 38% |
| **Lower-extremity bypass surgery** | | |
| Composite measure | 2.05 (1.42 to 2.95) | 32% |
| Risk-adjusted morbidity | 1.33 (0.91 to 1.93) | 3% |

**Figure 1.** Future risk-adjusted mortality rates (2009) for the "best" (top 20%), "middle," (middle 60%), and "worst" (bottom 20%) hospitals, as assessed using the composite measure and standard NSQIP techniques in the previous year (2008).

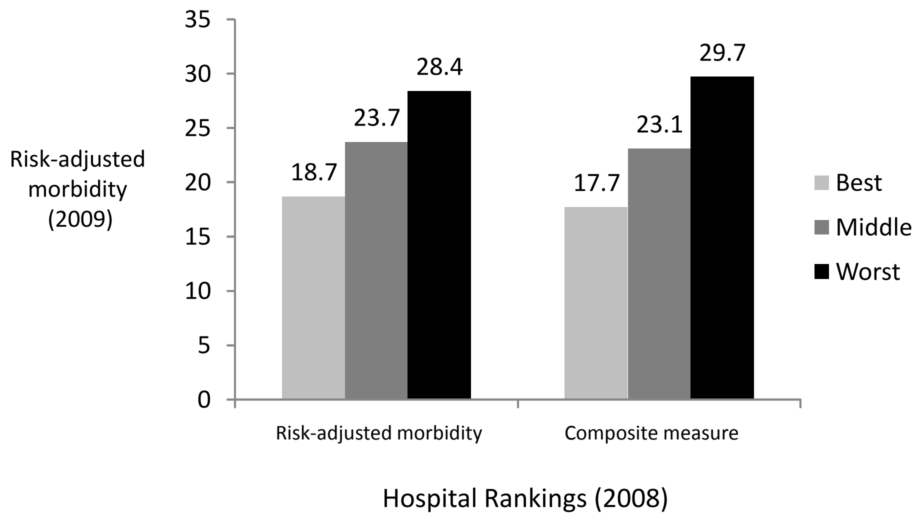
A. Ventral hernia repair



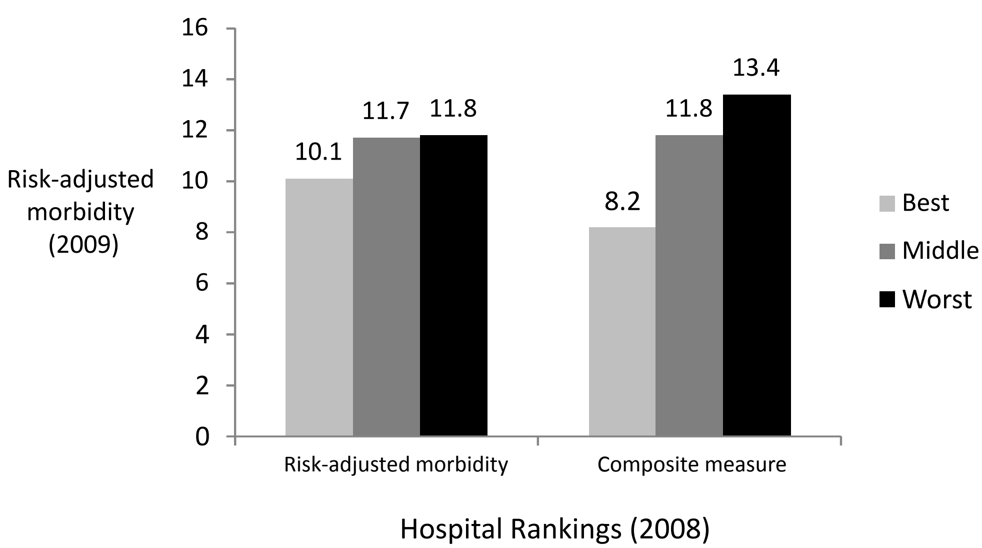

B. Colon resection



C. Elective abdominal aortic aneurysm repair




D. Lower-extremity bypass surgery



## Conclusions

In this study, we demonstrated the value of a novel composite measure for profiling hospitals on risk-adjusted rates of surgical morbidity. The Achilles heel of outcomes measurement is the unreliability due to small sample size and low event rates. As a result, the standard approach for assessing risk-adjusted morbidity and other surgical outcomes is prone to misclassification of surgeons and hospitals. The composite measure described in this paper addresses this problem in two ways: 1) applying statistical techniques for filtering out noise and 2) borrowing signals wherever they are available, including from other, related operations. In this study, we demonstrate that such a composite measure that integrates multiple outcomes, including morbidity with other, related procedures, is a better predictor of hospital performance than standard approaches for assessing risk-adjusted morbidity.

The findings of this study also demonstrate the value of incorporating information from other surgical procedures into a composite quality score. For each procedure, we found that adding risk-adjusted morbidity rates with “other” procedures enhanced the reliability of the hospital performance assessment. The ability to “borrow” signal from these other operations reflects the presence of a shared structure and process that leads to better outcomes for all surgical procedures, including nurse-to-patient ratios, quality improvement infrastructure, and adherence to evidence-based perioperative practices. Previous studies showing strong hospital-level correlations in surgical outcomes for different procedures (e.g., coronary artery bypass surgery and cardiac valve surgery) are consistent with these findings.

The results of this study should be viewed in the context of certain limitations. Because the ACS-NSQIP uses a sampling strategy (i.e., the registry does not capture 100% of cases in a hospital), our results may not be applicable to other quality measurement platforms. With 100% of the cases, the standard approach of assessing rates of hospital risk-adjusted morbidity would likely be more reliable. If morbidity rates were more reliable, the additional “signal” gained from other measures may not be as important. However, in our prior work we have seen the benefits of the composite measure persist in data sources that capture all patients. Moreover, this study is limited by the lack of information on structural characteristics, such as hospital volume. In our prior work, hospital volume is one of the most important inputs to the composite measure. If we added structural characteristics, the composite measure would likely be an even better predictor of future morbidity.

The Leapfrog Group, a large coalition of healthcare purchasers, now uses an approach analogous to the one in this paper for their evidence-based hospital referral program. This measure combines hospital mortality and provider volume into a single score that reflects the likelihood that a patient will survive surgery for five complex operations. These “Survival Predictor” scores are publicly reported on the Leapfrog Group website. These techniques were vetted and subsequently endorsed by the National Quality Forum for use with three high-risk surgical procedures: pancreatic resection, esophageal resection, and abdominal aortic aneurysm repair.

In addition to their value for public reporting, these composite measures could also be useful for quality improvement in the context of ACS-NSQIP and other reporting platforms. Standard approaches to surgical outcome measurements are plagued by statistical “noise” and imprecision, which translate into inaccurate assessments of relative hospital (or physician) performance. Such inaccurate assessments of performance can lead to both false positives (i.e., hospitals perceive a problem where none exist) and false negatives (i.e., hospitals miss a problem when it really does exist). The composite measures described in this paper could improve the reliability of benchmarking and give providers a truer sense of where they stand relative to their peers.

**List of Publications and Products**


1. **Dimick JB**, Staiger DO, Baser O, Birkmeyer JD. Composite measures for predicting surgical mortality in the hospital. Health Aff (Millwood) 2009;28:1189-98. PMID:19597221.
2. Ghaferi AA, Birkmeyer JD, **Dimick JB**. Variation in hospital mortality associated with inpatient surgery. N Engl J Med 2009;361:1368-75. PMID19797283.
3. **Dimick JB**, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? J Am Coll Surg 2010;210:503-8. PMC:2851222.
4. **Dimick JB**, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. Health Serv Res 2010;45:1614-29. PMC:2976775.
5. Osborne NH, Ko CY, Upchurch GR Jr, **Dimick JB**. Evaluating parsimonious risk-adjustment models for comparing hospital outcomes with vascular surgery. J Vasc Surg 2010;52:400-5. PMID:20670776.
6. Nicholas LH, Osborne NH, Birkmeyer JD, **Dimick JB**. Hospital process compliance and surgical outcomes in medicare beneficiaries. Arch Surg;145:999-1004. PMC2959198.
7. Osborne NH, Ko CY, Upchurch GR Jr, **Dimick JB**. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. J Vasc Surg 2011;53:1-5. PMID:21093202.
8. Ingraham AM, Cohen ME, Bilimoria KY, **Dimick JB**, Richards KE, Raval MV, Fleisher LA, Hall BL, Ko CY. Association of surgical care improvement project infection-related process measure compliance with risk-adjusted outcomes: implications for quality measurement. J Am Coll Surg 2010;211:705-14. PMID:21109157.
9. **Dimick JB**, Staiger DO, Osborne NH, Nicholas LH, Birkmeyer JD. Composite measures for rating hospital quality with major surgery. Health Serv Res 2012;47:1861-79. PMC:3448279.
10. Ryan A, Burgess J, Strawderman R, **Dimick JB**. What is the best way to estimate hospital quality outcomes? A simulation approach. Health Serv Res 2012;47:1699-718. PMID:22352894.

11. Kao LS, Ghaferi AA, Ko CY, **Dimick JB**. Reliability of superficial surgical site infections as a hospital quality measure. J Am Coll Surg 2011;213:231-5. PMC:3144290.
12. **Dimick JB**, Staiger DO, Hall BL, Ko CY, Birkmeyer JD. Composite measures for profiling hospitals on surgical morbidity. Ann Surg 2013;257(1):67-72. PMID:23235395.
13. Girotti ME, Ko CY, **Dimick JB**. Hospital morbidity rankings and complication severity in vascular surgery. J Vasc Surg 2013;57:158-64. PMC:3529764.
14. Hashmi ZG, **Dimick JB**, Efron DT, Haut ER, Schneider EB, Zafar SN, Schwartz D, Cornwell EE 3rd, Haider AH. Reliability adjustment: a necessity for trauma center ranking and benchmarking. J Trauma Acute Care Surg 2013;75:166-72. PMID:23940864.
15. Chen LM, Staiger DO, Birkmeyer JD, Ryan AM, Zhang W, **Dimick JB**. Composite quality measures for common inpatient medical conditions. Med Care 2013;51:832-837. PMID:23942222.
16. **Dimick JB**, Birkmeyer NJ, Finks JF, Share DA, English WJ, Carlin AM, Birkmeyer JD. Composite Measures for Profiling Hospitals on Bariatric Surgery Performance. JAMA Surg 2013 [Epub ahead of print]. PMID:24132708.