

# 1. Defining Quality in Surgery

*Justin B. Dimick*

## Introduction

With growing recognition of wide variations in surgical performance, demand for information on surgical quality is at an all time high. Patients and families are turning to their physicians, hospital report cards, and the Internet to identify the safest hospitals for surgery [1]. Payers and purchasers of health care are ramping up efforts to reward high quality (e.g., pay for performance) or steer patients toward the highest quality providers (e.g., selective referral) [2]. In addition to responding to these external demands, providers are becoming more involved in creating their own quality measurement platforms, such as the National Surgical Quality Improvement Program (NSQIP) [3]. Finally, professional organizations are now accrediting hospitals for some surgical services, including bariatric surgery [4].

Despite the need for good measures of quality in surgery, there is very little agreement about how to best assess surgical performance. According to the widely used Donabedian paradigm, quality can be measured using various aspects of structure, process, or outcome [5]. Recently, there is growing enthusiasm for composite, or “global,” measures of quality, which combine one or more elements of structure, process, and outcome [6]. In this chapter, we consider the advantages and disadvantages of each type of quality measure. We close by making recommendations for choosing among these different approaches.

## Structure

Structure refers to measurable attributes of a hospital (e.g., volume) or surgeon (e.g., specialty training) (Table 1.1). Because they are relatively easy to ascertain, measures of health care structure are widely

Table 1.1. Approaches to measuring the quality of care for aortic surgery with advantages and disadvantages of each approach.

Type of measure	Example	Advantages	Disadvantages
Structure	Hospital or surgeon volume	Inexpensive and readily available Good proxy for outcomes	Not actionable for quality improvement Not good for discriminating among individual providers
Process	Prophylactic antibiotics given on time Adherence to venous thromboembolism prevention guidelines	Actionable as targets for improvement Less influenced by patient risk and random errors	Known processes relate to unimportant or rare surgical outcomes Very few "high leverage" process of care are known
Outcome	Anastomotic leak rates with bariatric surgery Wound infection with ventral hernia repair	Seen as the bottom line of patient care Enjoy good "buy in" from surgeons	Sample sizes often too small at individual hospitals Need for detailed data for risk adjustment
Composite	Leapfrog Group's "Survival Predictor"	Addresses problems with small sample size Makes sense of multiple conflicting measures	Not granular enough to identify specific clinical areas that need improvement

used in health care. The American College of Surgeons (ACS) and the American Society of Metabolic and Bariatric Surgeons (ASMBS) are now accrediting hospitals for bariatric surgery based largely on measures of structure, including hospital volume, surgeon volume, and other structural elements necessary for providing multidisciplinary care for the morbidly obese [4].

Structural elements have several key strengths as quality measures. First, they are relatively easy to ascertain. Often, structural elements (e.g., volume) can be obtained from readily available administrative data. Second, many structural measures are strong predictors of hospital and surgeon outcomes. For example, with high-risk gastrointestinal surgery, such as pancreatic and esophageal resection, there are up to fivefold differences in mortality between high- and low-volume surgeons [7].

However, there are certain limitations of using structural quality measures. Most importantly, they are proxies for quality rather than direct measures. As a result, they only hold true on average. For example, while high-volume surgeons are better than low-volume surgeons on average, there are likely to be some high-volume surgeons with bad outcomes and low-volume surgeons with good outcomes [5]. Structural measures are also not actionable for quality improvement. Further, it is unclear how low-volume hospitals can change to replicate the excellent results of high-volume surgeons. Despite decades of research on the volume-outcome relationship, there is very little information about the details of care that differs between high-volume and low-volume hospitals [7].

## Process

Processes of care refer to those details of care that lead to good (or bad) outcomes. Using processes of care to measure quality is extremely common in ambulatory and inpatient medical care, but is not as widely used in surgery. Although processes of care in surgery can represent details of care in the preoperative, intraoperative, and postoperative phases of patient care, most existing process measures focus on details of preoperative patient care. For example, the Center for Medicare and Medicaid Services (CMS) Surgical Care Improvement Project (SCIP) measures focus on processes of care related to the prevention of complications, such as surgical site infection and venous thromboembolism.

Process measures have several strengths as quality measures (Table 1.1). First, processes of care are extremely actionable in quality

improvement. When hospitals and surgeon are “low outliers” for process compliance (e.g., patients not getting timely antibiotic prophylaxis), they know exactly where to target improvement. Second, in contrast to risk-adjusted outcomes measurement, processes of care do not need to be adjusted for differences in patient risk, which limits the need for data collection from the medical chart and saves valuable time and effort.

But using processes of care has several significant limitations in surgery. First, most existing process measures are not strongly related to important outcomes. For example, the SCIP measures, which are by far the most widely used process measure in surgery, are not related to surgical mortality, infections, or thromboembolism [8]. The lack of a relationship between SCIP measures and surgical mortality is easily explained by the fact that the complications they aim to prevent are secondary (e.g., superficial wound infection) or extremely rare (e.g., pulmonary embolism). However, there is also a very weak relationship between process measures and the outcome they are supposed to prevent (e.g., timely administration of prophylactic antibiotics and wound infection) [9]. This finding is more difficult to explain. It is possible that there are simply multiple other processes (many unmeasured or unmeasurable) that contribute to good surgical outcomes. As a result, it is likely that adherence to SCIP processes is necessary but not sufficient for good surgical outcomes.

## Outcome

Outcomes represent the end results of care. In surgery, the focus is often on operative mortality and morbidity. For example, the NSQIP, the largest clinical registry focusing on surgery, reports risk-adjusted morbidity and mortality rates to participating hospitals [3]. While morbidity and mortality have long been the “gold standard” in surgery, there is a growing focus on patient-oriented outcomes, such as functional status and quality of life.

Directly outcome measures have several strengths (Table 1.1). First, everyone agrees that outcomes are important. Measuring the end results of care makes intuitive sense to surgeons and other stakeholders. For example, the NSQIP has been enthusiastically championed by surgeons and other clinical leaders [10]. Second, outcomes feedback alone may improve quality. This so-called “Hawthorne effect” is seen whenever outcomes are measured and reported back to providers. For example, the

NSQIP in the Veterans Affairs (VA) hospitals and private sector has documented improvements over time that cannot be attributed to any specific efforts to improve outcomes [11].

However, outcome measures have key limitations. First, when the event rate is low (numerator) or the number of cases is small (denominator) outcomes cannot be reliably measured. Small sample size and low event rates conspire to limit the statistical power of hospital outcomes comparisons. For most operations, surgical mortality is too rare to be used as a reliable quality measure [12]. For example, a recent study evaluated seven operations for which mortality was advocated as a surgical quality measure by the Agency for Healthcare Research and Quality (AHRQ). The authors found that only one operation, coronary artery bypass surgery, had high enough caseloads to reliably measure quality with surgical mortality [13].

Another limitation of measuring outcomes is the need to collect detailed clinical data for risk adjustment [14]. Because patient differences can confound hospital quality measurement, it is important to adjust hospital comparisons for these differences in baseline risk. For example, the NSQIP presently collects more than 80 patient variables from the medical chart for this purpose [11]. This data collection is labor-intensive and expensive. Each NSQIP hospital employs a trained nurse clinician to collect this data.

## Composite

Composite measures are created by combining one or more structure, process, and outcome measures [6]. Composite measures offer several advantages over the individual measures discussed above (Table 1.1). By combining multiple measures, it is possible to overcome problems with small sample size discussed above. Composite measures also provide a “global” measure of quality. This type of measure is increasingly used for quality for value-based purchasing or other efforts that require an overall or summary measure of quality.

One key limitation with composite measures is that there is no “gold standard” approach for weighting input measures. Perhaps the most common approach is to weight each input measure equally. For example, in the ongoing Premier/CMS pay for performance demonstration project, Medicare payment bonuses are based on a composite score of process and outcome variables which are equally weighted. However,

this approach is severely flawed. Recent data show that variation in these composite measures is entirely driven by the process measures [15]. Newer approaches for empirically weighting individual measures will be discussed later.

Another limitation with composite measures is that they are not always actionable for quality improvement. By combining information on multiple measures and/or clinical conditions, there is often not enough “granularity” for clinicians to use the information for quality improvement. To target quality improvement efforts, it will often be necessary to deconstruct the composite into its component measures and find out where the problem lies (e.g., the specific procedure or complication).

## Choosing the Right Measurement Approach

No approach to quality measurement is perfect. Each type of measure – structure, process, and outcome – has its own strengths and limitations. In general, selecting the right approach to measure quality depends on characteristics of the procedure and the specific policy application [5].

Certain characteristics of the surgical procedure should be considered when selecting a quality measure (Fig. 1.1). Specifically, one should consider (1) how common adverse outcomes are and (2) how often an operation is performed. For procedures that are both common and relatively high risk (e.g., colectomy and gastric bypass), outcomes are reliable enough to be used as measures of quality (Fig. 1.1, Quadrant I). For procedures that are common but low risk (e.g., inguinal hernia repair), measures of process of care or functional outcomes are the best approach (Fig. 1.1, Quadrant II). For procedures that are high risk but uncommon (e.g., pancreatic and esophageal resection), structural measures such as hospital volume are likely the best approach (Fig. 1.1, Quadrant IV). In fact, empirical data suggests that structural measures such as hospital volume are better predictors of future performance than direct outcome measures for these uncommon, high-risk operations [16]. Finally, for operations that are both uncommon and low risk (e.g., Spigelian hernia repair), it is probably best to focus quality measurement efforts on other, more high leverage procedures.

When choosing an approach to quality measurement, the specific policy application should also be considered. In particular, it is important to distinguish between policy efforts aimed at selective referral and

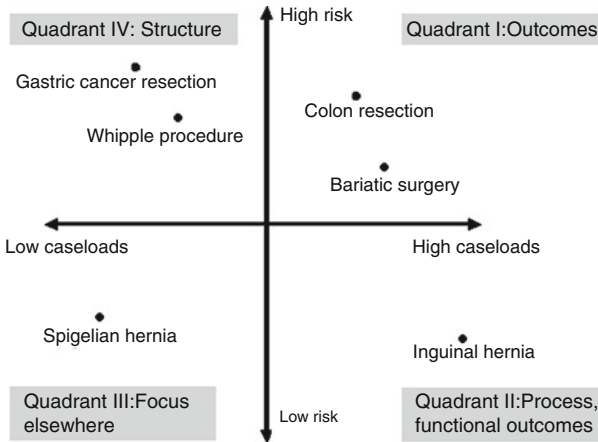


Fig. 1.1. Choosing among measures of structure, process, and outcomes. For high risk, high caseload operations (e.g., colectomy and bariatric procedures), outcomes are useful quality measures. For low risk, common procedures (e.g., inguinal hernia repair), processes of care or functional outcomes are appropriate measures. For high risk, uncommon operations (e.g., gastric and pancreatic cancer resection), measures of structure, such as hospital volume are most appropriate. For low risk, low caseload operations (e.g., spigelian hernia repair), it would be best to focus measurement efforts elsewhere. Figure modified by Birkmeyer et al. [5].

quality improvement. For selective referral, the main goal is to redirect patients to the highest quality providers. Structural measures, such as hospital volume, are particularly good for this purpose. Hospital volume tends to be strongly related to outcomes and large gains in outcomes could be achieved by concentrating patients in high-volume hospitals. In contrast, structural measures are not directly actionable and, therefore, do not make good measures for quality improvement. For improving quality, process, and outcome measures are better because they provide actionable targets. Surgeons and hospitals can improve by addressing problems with process compliance or focus on clinical areas with high rates of adverse outcomes. For example, the NSQIP reports risk-adjusted morbidity and mortality rates to every hospital. Surgeon champions and quality improvement personnel will target improvement efforts to areas where performance is statistically worse than expected.

## Improving Quality Measurement

Although the science of surgical quality measurement has come a long way in the past decade, it is still in its infancy. We will review several improvements to quality measurement currently on the horizon. These improvements focus on addressing the problems with the process of care and outcome measures discussed above.

We ultimately need to develop a better understanding of the processes of care that explain differences in outcome across hospitals. Once these “high leverage” processes of care are known, they can be promoted as best practices to improve care at all hospitals. Such research should use the tools of clinical epidemiology to isolate the root causes of variation in outcomes. For example, a recent study by Ghaferi and colleagues shed light on the mechanisms underlying variations in surgical mortality rates. Ghaferi et al., using detailed, clinically rich data from the NSQIP, ranked hospitals according to risk-adjusted mortality [17]. When comparing the “best” to “worst” hospitals, they found no significant differences in overall (24.6% vs. 26.9%) or major (18.2% vs. 16.2%) complication rates. However, the so-called “failure to rescue” (death following major complications) was almost twice as high in hospitals with very high mortality as in those with very low mortality (21.4% vs. 12.5%,  $p < 0.001$ ). This study highlights the need to focus on processes of care related to the timely recognition and management of complications – aimed at eliminating “failure to rescue” – to reduce variations in surgical mortality.

Recent emphasis has been placed on improving the efficiency of risk-adjustment techniques [18]. At present, most clinical registries collect a large number of clinical data elements from the medical record for risk adjustment. This “kitchen sink” approach to risk adjustment is largely based on the assumption that each additional variable improves our ability to make fair hospital comparisons. However, recent empiric data suggests that only the most important variables contribute meaningfully to risk-adjustment models. For example, Tu and colleagues demonstrated that a five-variable model provides nearly identical results to a 12-variable model for comparing hospital outcomes with cardiac surgery [19]. Using data from the NSQIP, we have demonstrated similar results for both general surgical procedures [18]. These results should be used to streamline the collection of data for risk adjustment, which will decrease the costs of data collection and lower the bar for participation in these important clinical registries.



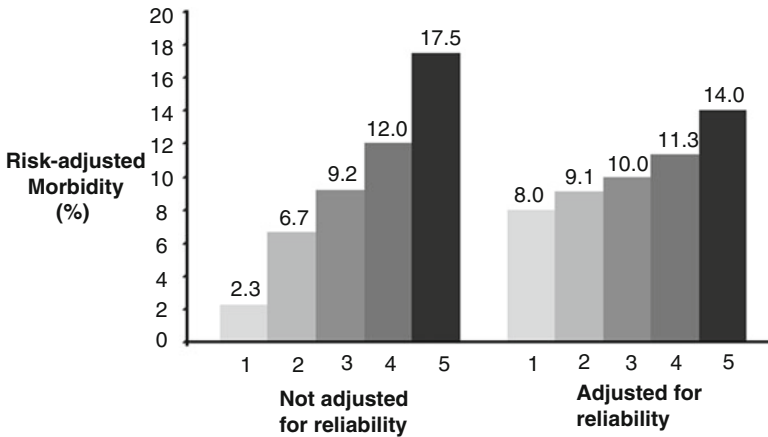


Fig. 1.2. Comparison of ventral hernia repair morbidity rates across hospital quintiles (1 = “best hospitals” and 5 = “worst hospitals”) before and after adjusting for statistical reliability. After adjusting for reliability, the apparent variation across hospitals is greatly diminished.

There is also increasing emphasis on using advanced statistical techniques for addressing the problem with “noisy” outcome measures [20]. As discussed above, imprecision from small sample size is the Achilles heel of outcomes measurement. These new techniques rely on empirical Bayes theory to adjust hospital outcomes for reliability. In this approach, the statistical “noise” is explicitly measured and removed by shrinking the observed outcome rate back toward the average rate. For example, Fig. 1.2 shows risk-adjusted hospital morbidity rates across quintiles for ventral hernia repair, before and after adjusting for reliability. Before adjusting for reliability, rates of morbidity varied eightfold (2.3–17.5%) from the “best” to “worst” quintile. However, after removing chance variation (i.e., “noise”) by adjusting for reliability, rates of morbidity varied less than twofold (8.0–14.0%) from the “best” to “worst” quintile.

While this approach has many advantages, reliability adjustment makes the assumption that small hospitals have average performance. Although this approach gives small hospitals, the benefit of the doubt (i.e., they are innocent until proven guilty), under certain circumstances it could bias hospital rankings. For instance, given the well-known relationship between volume and outcome in surgery, these small hospitals may actually have performance below average. Incorporating

information about hospital volume could address this bias. We have developed a novel technique for performing reliability adjustment by shrinking to a conditional average (i.e., the outcome expected given hospital volume) to address this problem [6]. This approach is considered a composite measure as it includes two inputs (mortality and volume).

This general approach can also be used to create more sophisticated composite measures of quality. As discussed above, most current approaches for combining measures are flawed. To address this problem, we have developed a method for empirically weighting input measures [21]. Briefly, we first identify a gold standard quality measure, such as mortality or serious morbidity. We then determine the relationship between each candidate measure and this gold standard measure. Finally, each input measure is given a weight based on (1) the reliability with which it is measured and (2) how correlated it is with the gold standard measure. These empirically weighted composite measures been shown to be better predictors of future performance than individual measures alone [21].

## Conclusions

Each type of quality measure – structure, process, and outcome – has its unique strengths and limitations. Structural measures are strongly related to important outcomes and are readily available. Unfortunately, however, structural measures are proxies for quality and do not discriminate among individual providers. Process measures are extremely useful because they are actionable for quality improvement. But the most high leverage processes in surgery are not yet known. Outcomes are the bottom line in surgery and everyone agrees that they are important. Because of small sample size at most hospitals, however, they are often too “noisy” to reliably reflect hospital quality. Ultimately, when choosing among these different approaches, surgeons need to be flexible and consider the specific procedure and policy application prior to choosing a measure.

## Selected Readings

1. Osborne NH, Nicholas LH, Ghaferi AA, et al. Do popular media and internet-based hospital quality ratings identify hospitals with better cardiovascular surgery outcomes? *J Am Coll Surg.* 2010;210:87–92.

2. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA*. 2007;297:740–4.
3. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg*. 2008;207:777–82.
4. Dimick JB, Osborne NH, Nicholas L, et al. Identifying high-quality bariatric surgery centers: hospital volume or risk-adjusted outcomes? *J Am Coll Surg*. 2009;209:702–6.
5. Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? *J Am Coll Surg*. 2004;198:626–32.
6. Dimick JB, Staiger DO, Baser O, et al. Composite measures for predicting surgical mortality in the hospital. *Health Aff (Millwood)*. 2009;28:1189–98.
7. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–37.
8. Hawn MT. Surgical care improvement: should performance measures have performance measures. *JAMA*. 2010;303:2527–8.
9. Stulberg JJ, Delaney CP, Neuhauser DV, et al. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA*. 2010;303:2479–85.
10. Neuman HB, Michelassi F, Turner JW, et al. Surrounded by quality metrics: what do surgeons think of ACS-NSQIP? *Surgery*. 2009;145:27–33.
11. Khuri SF, Daley J, Henderson WG. The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. *Arch Surg*. 2002;137:20–7.
12. Dimick JB, Welch HG. The zero mortality paradox in surgery. *J Am Coll Surg*. 2008;206:13–6.
13. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA*. 2004;292:847–51.
14. Iezzoni LI. The risks of risk adjustment. *JAMA*. 1997;278:1600–7.
15. O'Brien SM, DeLong ER, Dokholyan RS, et al. Exploring the behavior of hospital composite performance measures: an example from coronary artery bypass surgery. *Circulation*. 2007;116:2969–75.
16. Birkmeyer JD, Dimick JB, Staiger DO. Operative mortality and procedure volume as predictors of subsequent hospital performance. *Ann Surg*. 2006;243:411–7.
17. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009;361:1368–75.
18. Dimick JB, Osborne NH, Hall BL, et al. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg*. 2010;210:503–8.
19. Tu JV, Sykora K, Naylor CD. Assessing the outcomes of coronary artery bypass graft surgery: how many risk factors are enough? Steering Committee of the Cardiac Care Network of Ontario. *J Am Coll Cardiol*. 1997;30:1317–23.
20. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res*. 2010;45:1614–29.
21. Staiger DO, Dimick JB, Baser O, et al. Empirically derived composite measures of surgical performance. *Med Care*. 2009;47:226–33.