

# Measuring the Quality of Surgical Care: Structure, Process, or Outcomes?

John D Birkmeyer, MD, FACS, Justin B Dimick, MD, Nancy JO Birkmeyer, PhD

With widespread recognition that surgical outcomes vary by provider,<sup>1,2</sup> surgeons and hospitals are increasingly being asked to provide evidence of the quality of care that they deliver. Patients and their families are turning to the Internet and other sources to make better informed decisions about where and by whom to undergo surgery. Both public and private payers are looking to steer selected populations of surgical patients to high-quality providers—so-called value-based purchasing.<sup>3</sup> To meet these interests, policy makers, health services researchers, and a variety of related organizations have redoubled their efforts to develop and implement quality indicators germane to surgery.

There remains considerable debate about which measures should be used to reflect surgical quality. Structural measures—a very broad group of variables that reflect the setting in which care is delivered—have received considerable attention lately. For example, the Leapfrog Group, a large coalition of health-care purchasers, is encouraging patients to seek care at hospitals with high procedure volumes for several operations.<sup>4</sup> Process measures, which reflect the particulars of care that patients actually receive, have long served as quality indicators in primary care and other specialties (eg, use of  $\beta$ -blockers after myocardial infarction). There is evidence that focusing on process measures may be equally useful in surgery.<sup>5</sup> Finally, and most obviously, the quality of surgical care can be assessed by direct outcomes measurement. Quality improvement programs focusing on risk-adjusted morbidity and mortality rates have long been

standard in cardiac surgery and in hospitals of the Department of Veterans Affairs.<sup>6-8</sup>

In this article, we consider the relative merits of these different approaches to measuring and ultimately improving the quality of surgical care. Adopting the Donabedian paradigm,<sup>9</sup> we consider quality measurement in three domains: structure, process, and outcomes. Although each of these three approaches has unique advantages, each has its own conceptual and practical limitations (Table 1).

## Structural measures

Structural measures include a broad list of variables reflecting the setting or system in which care is delivered. These may describe hospital's physical plant and resources. They also include measures that relate directly or indirectly to staff expertise or staff coordination and organization. Of these variables, procedure volume, measured at either the surgeon or hospital level, is most commonly used as a surrogate for surgical quality. Although the magnitude of volume-outcomes associations with various procedures is debated, there is little doubt that high-volume providers have lower operative mortality, fewer complications, or better longterm survival with some operations than their lower-volume counterparts.<sup>10-12</sup>

Among other structural variables, subspecialty training by the operating surgeon is often cited as a predictor of improved surgical outcomes. For example, patients undergoing resection for rectal cancer had lower recurrence rates and improved survival when treated by surgeons board certified in colorectal surgery.<sup>13</sup> Structural variables more broadly related to staff organization and resource availability may also influence surgical outcomes. For example, a considerable body of evidence has accrued suggesting that critically ill surgical patients have lower mortality in "closed" intensive care units—those in which patients are managed primarily by dedicated, board-certified intensivists.<sup>14</sup> Similarly, hospitals with high nurse-to-bed ratios seem to have lower mortality rates for some operations.<sup>15</sup> Finally, resource avail-

### No competing interests declared.

Supported by a grant from the National Cancer Institute (1 RO1 CA098481-01A1). Dr John Birkmeyer is a consultant for the Leapfrog Group and chairs its expert panel on Evidence-Based Hospital Referral. Dr Dimick is also a consultant for the Leapfrog Group.

Received September 17, 2003; Accepted November 26, 2003.

From the Center for Surgical Evaluation and Policy, Department of Surgery (Dimick), University of Michigan, Ann Arbor, MI, and VA Outcomes Group, VA Medical Center (Birkmeyer, Dimick, Birkmeyer), White River Junction, VT.

Correspondence address: John D Birkmeyer, Section of General Surgery, University of Michigan, 2920 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0331.

**Abbreviations and Acronyms**

CABG = coronary artery bypass grafting  
 NSQIP = National Surgical Quality Improvement Program

ability may be an important determinant of surgical outcomes. For example, one study from the Department of Veterans Affairs found that hospitals with lower than expected mortality rates tended to have more up-to-date technology and equipment in their intensive care units.<sup>16</sup>

**Advantages of structural variables**

From a measurement perspective, structural measures have several attractive features as indicators of surgical quality. As already described, many of these variables are strongly related to surgical outcomes. For example, with esophagectomy and pancreatic resection, operative mortality rates at very high volume hospitals are on average 10% lower, in absolute terms, than at lower-volume centers. The primary advantage of structural variables is expediency.<sup>17</sup> Compared with direct outcomes assessment, structural variables, including procedure volume,

can be assessed easily and inexpensively, often with administrative data.

**Disadvantages of structural variables**

Among the downsides, the literature assessing structural measures is incomplete. It focuses on a small number of variables (eg, volume) and outcomes measures (eg, operative mortality). Little is known about the importance of structural variables that are more difficult to measure or about relationships between structure and nonfatal outcomes. Unlike process measures, which can often be evaluated in randomized clinical trials, most structural measures can only be assessed in observational studies. It is often difficult to rule out confounding as an explanation for observed associations between structure and outcomes. Second, in contrast to process measures, many structural measures are not readily actionable, which limit their ultimate effectiveness as a means toward quality improvement. For example, a small hospital can increase how many of its high-risk patients receive perioperative  $\beta$ -blockers, but it cannot readily make itself a high-volume center for a given procedure or, unless it has sufficient staff, convert to a closed-model intensive care unit.

**Table 1.** Using Structure, Process, and Outcomes to Measure Surgical Quality, with Examples, Advantages, and Disadvantages of Each

	<b>Structure</b>	<b>Process</b>	<b>Outcomes</b>
Examples	Procedure volume	Perioperative $\beta$ -blockers in high-risk surgical patients	Morbidity and mortality rates
	Fellowship-trained surgeons	Use of internal mammary graft during coronary artery bypass graft	Functional health status
	“Closed” intensive care units		Patient satisfaction Cost
Primary advantage(s)	Expedient, inexpensive proxies of surgical outcomes	Reflect care that patients actually receive—may seem “fairer” to providers	Buy-in from surgeons—the “bottom line” of what they do
		Actionable from provider perspective, clear link to quality improvement activities	Outcomes measurement alone may improve outcomes
Disadvantages	Most variables not actionable from provider perspective	Little information about which processes are important for specific procedures	Numbers too small to measure with adequate precision procedure-specific outcomes for most hospitals and procedures
	Imperfect proxies for outcomes—reflect average results for large groups of providers, not individuals		Outcomes measures that are not procedure-specific less useful for purposes of quality improvement

**Table 2.** Examples of Process Measures Associated with Surgical Outcomes, According to Strength of Scientific Evidence Supporting Them and the Cost and Complexity of Implementing Them

Implementation cost/complexity	Strength of the evidence/magnitude of potential benefit				
	Greatest	High	Medium	Lower	Lowest
Low	Appropriate venous thromboembolism prophylaxis	Use of supplemental perioperative oxygen	H <sub>2</sub> antagonists for stress ulcer prophylaxis	Early analgesics in patients with acute abdomen (without compromising diagnostic accuracy)	Use of pre-anesthesia checklists
	Use of perioperative $\beta$ -blockers in patients at risk for cardiac events	Selective decontamination of digestive tract	Protocols for intravenous heparin titration	Intraoperative monitoring of vital signs/oxygenation	Counting sharps, instruments, and sponges
	Appropriate use of antibiotic wound prophylaxis Antibiotic-impregnated central venous catheters	Use of silver alloy-coated urinary catheters	Maintenance of perioperative normothermia		Sucralfate for stress ulcer prophylaxis
Medium or high	Early nutritional support in critically ill patients		Barrier precautions (via gowns and gloves; dedicated equipment and personnel)	Tunneling short-term central venous catheters	“Sign your site” protocols
	Use of real-time ultrasound guidance during central line insertion		Perioperative glucose control		Catheter changes as prophylaxis against central line infections

Based on literature review and conclusions contained in “Making Health Care Safer: A Critical Analysis of Patient Safety Procedures,” from the Agency for Healthcare Quality and Research.<sup>18</sup>

Finally, and most importantly, structural variables are very imperfect proxies for quality—they reflect average results for large groups of providers, not individuals. For example, many low-volume hospitals have excellent performance, but many high-volume centers are poor performers. Even if all high-risk procedures were concentrated in high-volume hospitals, there would remain substantial variation in quality across hospitals and this opportunity for improvement.

### Process measures

Process variables describe the care that patients actually receive and are routinely used as quality indicators in nonsurgical specialties. For example, in large managed care organizations and Department of Veterans Affairs hospitals, primary care physicians are regularly graded according to the proportion of appropriate patients in their practices who receive screening mammography, retinal examinations (in diabetics), or pneumococcus vaccinations. Similarly, providers are assessed in terms of the proportion of patients surviving a myocardial infarction who are discharged on aspirin and  $\beta$ -blockers.

Although not yet used widely as quality indicators in

surgery, many processes of care are strongly associated with improved patient outcomes. For example, the Agency for Healthcare Research and Quality recently commissioned a critical review of a large number of hospital-based practices related to patient safety (Table 2).<sup>18</sup> A large number of practices related to perioperative care have high levels of evidence supporting their effectiveness. These include practices related to central venous line management, critical care, and minimizing risks of postoperative cardiac events, venous thromboembolism, and other complications.

Procedure-specific processes of care may sometimes explain apparent associations between structural variables and outcomes. For example, Hannan and colleagues performed a prospective clinical study of patients undergoing carotid endarterectomy at six hospitals in New York State.<sup>19</sup> In that study, vascular surgeons had substantially lower 30-day rates of operative stroke or death than did general surgeons or neurovascular surgeons. The investigators also found that use of intraarterial shunting, eversion endarterectomy techniques, patching of the arteriotomy, and protamine were associated with lower complication

rates. Greater adoption of these four processes of care by vascular surgeons explained in large part their better outcomes.

### **Advantages of process measures**

As potential quality indicators, process of care measures have several attractive features. In addition to the high level of evidence supporting their effectiveness (often randomized clinical trials), some process measures have very large potential benefits. For example, in one large trial, patients receiving  $\beta$ -blockers during and after major noncardiac surgery had much lower 1-year mortality than patients who did not (3% versus 14%,  $p < 0.005$ ).<sup>20</sup> Second, process variables reflect the care that patients actually receive and may be perceived by providers as “fairer” measures of quality than structural measures.

Finally, and most importantly, process of care measures are generally actionable and link directly to quality improvement activities. For example, investigators and clinicians at six hospitals in northern New England have maintained a prospective clinical registry for coronary artery bypass graft (CABG) and other cardiac procedures since 1987. They identified numerous process of care measures linked to lower operative mortality, including use of an internal mammary graft, continuing aspirin through surgery, and maintaining a hematocrit of 24% or higher when “on pump.” As a result of systematic efforts to increase the use of these practices and timely feedback of performance data to clinicians, operative mortality rates across the region fell by almost half during the 1990s, a decline significantly greater than observed in regions of the United States without similar quality improvement initiatives in place.<sup>21</sup>

### **Disadvantages of process measures**

Measurement systems focusing on process variables must be able to accurately identify eligible patient populations (ie, the right denominator). Many processes known to be effective in general may not be appropriate for all patients undergoing a given procedure (eg,  $\beta$ -blockers in patients with bradyarrhythmias or severe left ventricular dysfunction). Ensuring the right denominator implies the need for clinical data and may be labor intensive, a practical limitation of process measurement.

A second major limitation of process measures is the relative lack of evidence about which processes are important for specific procedures. Much of the existing literature on processes of care focus on the medical management of surgical patients (Table 2). Many of the most

serious adverse events occurring after surgery are non-medical in nature, arising from technical problems associated with the procedure itself—anastomotic leaks, bleeding, or wound complications. Although high-leverage technical processes have been elucidated for some procedures (notably CABG and carotid endarterectomy), few procedures have been as carefully studied, leaving major knowledge gaps.

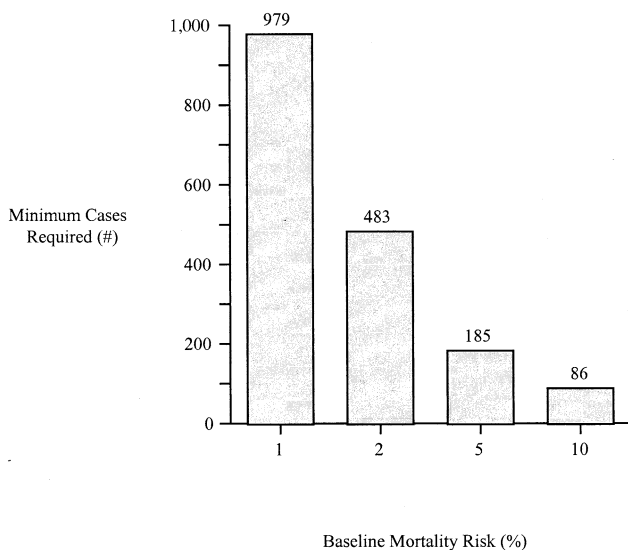
### **Direct outcomes measures**

Since surgeon Ernest A Codman began tracking the “end results” of surgical procedures in the early 20<sup>th</sup> century,<sup>22</sup> direct outcomes assessment has long been a staple in assessing the quality of surgical care. Although operative mortality is most commonly used, other outcomes measures that could be considered quality indicators include complication rates, length of stay, readmission rates, patient satisfaction, functional health status, and other measures of health-related quality of life.

There are many ongoing, large-scale initiatives aimed specifically at measuring and improving surgical outcomes. Clinical outcomes registries in cardiac surgery, including those launched in New York, Pennsylvania, and northern New England in the 1980s, were among the earliest and most successful.<sup>8,23</sup> More states and regions and one national organization (the Society for Thoracic Surgeons) have since implemented similar data collection systems. Although these registries vary in many respects, all provide hospitals and cardiac surgeons with feedback on their risk-adjusted morbidity and mortality rates. Over the past decade, prospective outcomes registries have been implemented in numerous other fields. Although most outcomes measurement efforts have been procedure-specific, the National Surgical Quality Improvement Program (NSQIP) of the Department of Veterans Affairs assesses hospital-specific morbidity and mortality rates aggregated across a wide range of surgical specialties and procedures. Efforts to apply the same measurement approach outside the Veterans Affairs are currently under way.<sup>24</sup>

### **Advantages**

Direct outcomes measures have at least two major advantages. First, because most consider patient outcomes the “bottom line” of surgical practice, efforts assessing quality with direct outcomes measures have obvious face validity and are likely to get the greatest buy-in from surgeons. Second, measurement alone may improve



**Figure 1.** Minimum number of cases required at individual hospitals to identify a statistically significant doubling of baseline mortality risk.

outcomes—the so-called Hawthorne effect. Surgical morbidity and mortality rates in Veterans Affairs hospitals have fallen dramatically since implementation of NSQIP in 1991.<sup>25</sup> No doubt many surgical leaders at individual hospitals made specific organizational or process improvements after they began receiving feedback on their hospitals' performance. It is very unlikely that even a full inventory of these specific changes would explain such broad-based and substantial improvements in morbidity and mortality rates.

### Disadvantages

The most important limitation of direct outcomes measurement relates to sample size. For the large majority of surgical procedures, very few hospitals (or surgeons) have sufficient adverse events (numerators) and cases (denominators) for meaningful, procedure-specific measures of morbidity or mortality. Consider a hypothetical hospital with an observed mortality rate of 10% for a given procedure, twice the national average of 5%. That hospital's rate would need to be based on at least 185 cases to be reasonably confident (95%) that its performance was significantly worse than the national average and not simply indicative of random variation (Fig. 1). Outside of cardiac surgery, very few procedures have baseline mortality rates of 5% or higher and are performed this frequently at individual hospitals (particularly low-volume ones). Most common operations tend to be associated with low baseline risks, which substan-

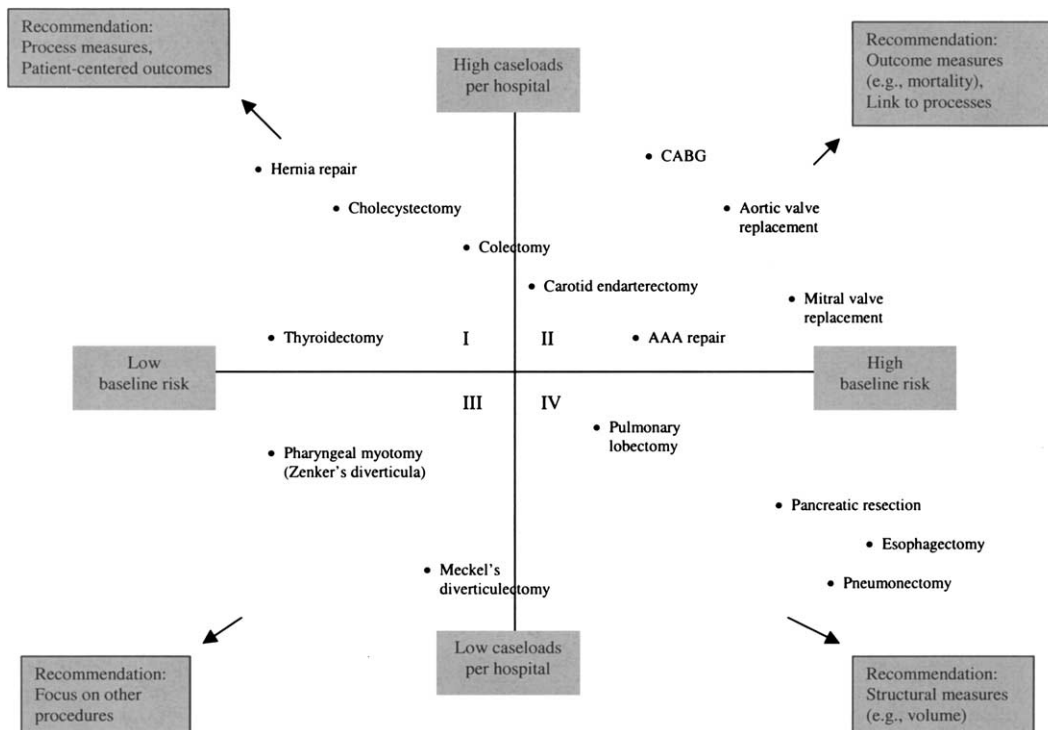
tially compounds problems with statistical power in measuring outcomes at the provider level (Fig. 1).

To circumvent sample size limitations with procedure-specific measures, hospitals and surgeons could determine morbidity and mortality rates after aggregating a wide range of procedures across different surgical specialties (eg, the NSQIP approach). Unfortunately, this approach is less satisfying from a quality improvement perspective. Understanding and improving the delivery of a specific procedure may require measures specific to that operation. Aggregated measures of surgical morbidity and mortality may also be suboptimal for patients (and payers) interested in identifying excellence with individual operations. In other words, measures weighted heavily toward outcomes of common, low-risk operations (eg, hernia repairs, cholecystectomies) may not be very informative for patients deciding where to undergo a Whipple operation.

### Choosing the right measure

Although structural, process, and outcomes measures all have unique strengths, these three measures have distinct downsides, depending on how they are used. For these reasons, both surgeons and policy makers should be flexible in their approach to measuring quality and develop strategies best suited to meeting specific needs.

The procedure itself may be the most important factor in deciding about the most effective approach to quality measurement. Two attributes are particularly important: 1) the baseline risks of the procedure and 2) how commonly it is performed at individual hospitals (Fig. 2). Measuring quality for procedures that are both low risk and uncommonly performed (Quadrant III) should receive low priority. Many high-risk procedures, such as esophagectomy and pancreatic resection (Quadrant IV), are performed too infrequently at the vast majority of hospitals to support direct outcomes assessment. Procedure volume, a structural measure highly correlated with mortality for many of these procedures, is likely the only practical quality indicator. Quality for procedures that are both common and relatively high risk (eg, CABG, Quadrant II) is best assessed directly using risk-adjusted measures of morbidity and mortality. Quality improvement consortia designed to accomplish this task are also ideal platforms for measuring process variables and linking them to outcomes. Measuring quality is perhaps most problematic for common but relatively low-risk procedures (eg, laparoscopic cholecystectomy, Quadrant I). For these procedures, volume and



**Figure 2.** Recommendations for when to focus on structure, process, or outcomes.

other structural measures are not known to be major determinants of outcomes. Low baseline rates of mortality or other serious complications preclude measuring outcomes with sufficient precision. Quality for these procedures is best judged by process measures (where available) or by outcomes measures other than morbidity and mortality (eg, functional health status).

The right measure also depends on the specific policy context and the ultimate goal of quality measurement. With public reporting initiatives, for example, the primary goal is to inform patients about where to undergo elective surgical procedures. In this context, quality measures must have strong face validity for patients. Despite its many limitations, procedure volume, a structural measure, is considered important by many patients, who frequently ask their surgeons, "Do you do this procedure often?" For purposes of quality improvement, quality measures should be selected primarily on the basis of their validity as judged by providers and their relative actionability. Quality improvement requires rigorous measurement of process and outcomes.

A final, practical consideration is the potentially high cost of quality assessment. Though information about structural variables can be obtained expediently using existing data, process and outcomes measures imply the

need for clinical data collection systems, which can be expensive. For example, participation in NSQIP costs approximately \$40 per operation.<sup>26</sup> Hospitals, if not payers, need to be prepared for this level of investment if they are to pursue data-driven quality improvement initiatives.

### Improving the quality of quality measurement

As this article suggests, current tools for measuring surgical quality are far from perfect. Opportunities for moving the field forward by focusing exclusively on structural measures (proxy variables like volume) are limited. Improving the quality of quality measurement will require progress in other areas. Identifying high-leverage processes of care is clearly one of them. As described earlier, most high-level evidence linking process to surgical outcomes pertains to the medical aspects of perioperative care, not the technical aspects of specific procedures that determine their success. A better understanding of such processes is essential if successes achieved with CABG are to be replicated in other areas.

Second, we should place a higher priority on measuring patient-centered outcomes. To date, most large-scale quality improvement initiatives in surgery have focused on measures of morbidity and mortality. Though such

outcomes may be central for many cardiovascular and cancer procedures, they are considerably less useful for assessing the quality of low-risk operations, particularly those whose primary goal is improving health-related quality of life. Valid and reliable instruments for assessing health-related quality of life are widely available.<sup>27</sup> The more difficult task is finding ways to collect such data efficiently and inexpensively.

Finally, we must be careful to avoid missing the big picture. Quality in health care can be described as “doing the right things right.” This article (and debates about surgical quality in general) focuses on only the latter component of this aphorism—how well the procedure was performed (ie, doing things right). As suggested by wide geographic variation in the use of different procedures in the United States, the quality of the decision to operate in the first place (ie, doing the right thing) may be an equally important issue in surgical care. A full accounting of surgical quality will require measures of appropriateness and how well patient preferences are incorporated in clinical decisions, in addition to those assessing how well they do after surgery.

#### Author Contributions

Study conception and design: J Birkmeyer, Dimick, N Birkmeyer

Drafting of manuscript: J Birkmeyer

Critical revision: Dimick, N Birkmeyer

Obtaining funding: J Birkmeyer

#### REFERENCES

1. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997;185:315–327.
2. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. *JAMA* 1991;266:803–809.
3. Galvin R, Milstein A. Large employers' new strategies in health care. *N Engl J Med* 2002;347:939–942.
4. Birkmeyer JD, Wennberg DE, Young M, Birkmeyer CB. Leapfrog safety standards: potential benefits of universal adoption. Washington, DC: The Business Roundtable; 2000.
5. Ferguson TB Jr, Peterson ED, Coombs LP, et al. Use of continuous quality improvement to increase use of process measures in patients undergoing coronary artery bypass graft surgery: a randomized controlled trial. *JAMA* 2003;290:49–56.
6. Hannan EL, Kilburn H Jr, Racz M, et al. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;271:761–766.
7. Khuri SF, Daley J, Henderson W. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg* 1998;228:491–507.
8. O'Connor GT, Plume SK, Morton JR, et al. Results of a regional prospective study to improve the in-hospital mortality associated with coronary artery bypass grafting. *JAMA* 1996;275:841–846.
9. Donabedian A. Evaluating the quality of medical care. *Milbank Memorial Fund Q* 1966;44:166–206.
10. Dudley RA, Johansen KL, Brand R, et al. Selective referral to high volume hospitals: estimating potentially avoidable deaths. *JAMA* 2000;283:1159–1166.
11. Begg CB, Reidel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med* 2002;346:1138–1144.
12. Finlayson EV, Birkmeyer JD. Effects of hospital volume on life expectancy after selected cancer operations in older adults: a decision analysis. *J Am Coll Surg* 2003;196:410–417.
13. Porter GA, Soskolne CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. *Ann Surg* 1998;227:157–167.
14. 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.
15. Needleman J, Buerhaus P, Mattke S, et al. Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med* 2002;346:1715–1722.
16. Daley J, Forbes MG, Young GJ, et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. *J Am Coll Surg* 1997;185:341–351.
17. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002;346:1128–1137.
18. Making health care safer: A critical analysis of patient safety practices. Rockville, MD: Agency for Healthcare Research and Quality; 2001.
19. Hannan EL, Popp AJ, Feustel P, et al. Association of surgical specialty and processes of care with patient outcomes for carotid endarterectomy. *Stroke* 2001;32:2890–2897.
20. Mangano DT, Layug EL, Wallace A, Tateo I. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. Multicenter Study of Perioperative Ischemia Research Group. *N Engl J Med* 1996;335:1713–1720.
21. Peterson ED, DeLong ER, Jollis JG, et al. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *J Am Coll Cardiol* 1998;32:993–999.
22. Neuhauser D. Ernest Amory Codman MD. *Qual Health Care* 2002;11:104–105.
23. Hannan EL, Kilburn H Jr, O'Donnell JF, et al. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. *JAMA* 1990;264:2768–2774.
24. Fink AS, Campbell DA Jr, Mentzer RM Jr, et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Ann Surg* 2002;236:344–353.
25. 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;137:20–27.
26. Khuri SF. Quality, advocacy, healthcare policy, and the surgeon. *Ann Thor Surg* 2002;74:641–649.
27. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press; 1996.