

Healthcare Quality Measurement in Orthopaedic Surgery

Current State of the Art

Andrew Auerbach MD, MPH

Published online: 21 April 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Improving quality of care in arthroplasty is of increasing importance to payors, hospitals, surgeons, and patients. Efforts to compel improvement have traditionally focused measurement and reporting of data describing structural factors, care processes (or ‘quality measures’), and clinical outcomes. Reporting structural measures (eg, surgical case volume) has been used with varying degrees of success. Care process measures, exemplified by initiatives such as the Surgical Care Improvement Project measures, are chosen based on the strength of randomized trial evidence linking the process to improved outcomes. However, evidence linking improved performance on Surgical Care Improvement Project measures with improved outcomes is limited. Outcome measures in surgery are of increasing importance as an approach to compel care improvement with prominent examples represented by the National Surgical Quality Improvement Project. Although outcomes-focused approaches are often costly, when linked to active benchmarking and collaborative activities, they may improve care broadly. Moreover, implementation of computerized data systems collecting information formerly collected on paper only will facilitate benchmarking. In the end, care will only be improved if

these data are used to define methods for innovating care systems that deliver better outcomes at lower or equivalent costs.

Introduction

Healthcare quality and safety improvement is a national imperative with reports such as the Institute of Medicine’s “To Err Is Human” and “Crossing the Quality Chasm” providing the catalyst to compel broad system change. In the nearly 8 years since these reports’ publication, a number of national initiatives seek to measure care systems so that care structures (eg, the sociologic or infrastructural systems), care processes (eg, treatments provided to patients), and our patient outcomes can be collected and fairly compared across sites. These data can then be used by caregivers to compel change within their site, by purchasers seeking to choose preferred providers, by regulators with the goal of accrediting hospitals, and by patients making an informed choice about where to seek care.

Understanding how to compare these elements of care—structures, processes, and outcomes—across providers is of importance in arthroplasty because total joint arthroplasty is a highly common and costly procedure, one that is highly effective at returning function and improving quality of life. In this context, unexpected or unnecessary complications become even more egregious lapses in care.

We review an expanded version of a classic model for assessing healthcare quality. We use the expanded structure + process = outcome model to provide a framework within which we review examples of structure-focused initiatives to improve care, process (or quality-measure)-focused initiatives, and outcome-driven

The author (AA) has received research funding from the National Heart Lung and Blood Institute, Agency for Research and Quality, and California Healthcare Foundation.

A. Auerbach (✉)
UCSF Division of Hospital Medicine, 533 Parnassus Avenue,
Room U131, UCSF Box 0131, San Francisco, CA 94143-0131,
USA
e-mail: Ada@medicine.ucsf.edu

collaboratives, which hope to improve patient care, although via different mechanisms.

improve care by targeting one or more of the constituent parts.

A Conceptual Model of Factors Associated With Healthcare Outcomes

Donabedian first proposed a highly useful model outlining factors contributing to patient outcomes in 1966 (Fig. 1) [10]. This model posits that care structures (such as having a dedicated arthroplasty ward or care team) and care processes (such as a standard arthroplasty care pathway) can contribute to patient outcomes. Outcomes can include patient-centered experiences, use of resources during the episode of care (such as costs), clinical events such as mortality, major complications, readmission, functional status, pain, and ability to return to work.

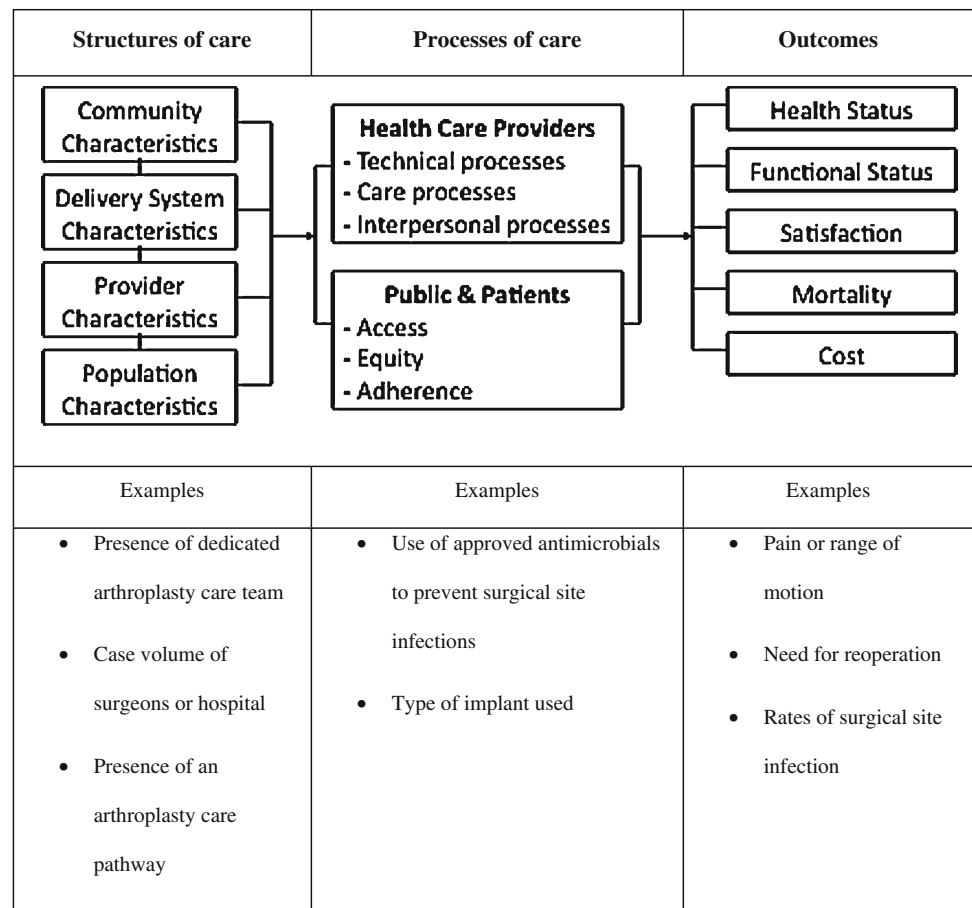
Understanding this model is relevant to health system measurement in that it also provides the framework within which one can identify and specify measures relevant to one's practice. In fact, many of these domains have been the subject of national and regional initiatives that seek to

Structural Measurement as a Way to Compel Change

The Leapfrog Group [2] is a prominent example of a healthcare system improvement initiative that, at least in its early phases, primarily targeted care structures as a way to compel change [2]. Founded in 1998 by a consortium of large purchasers, the Leapfrog Group specified three major structural measures in its early iterations: (1) adoption of computerized physician order entry (CPOE); (2) use of critical care board-certified physicians in intensive care units; and (3) adherence to volume standards for selected high-risk surgical procedures.

These measures were selected based in large part on the evidence associating them with improved outcomes for hospitalized patients. For example, volume measures for high-risk surgery (which would imply that certain surgeries would need to be regionalized) would result in a substantial number of deaths being averted; adherence to all Leapfrog measures would provide even more benefits [2]. Structural

Fig. 1 The Donabedian model of how healthcare system factors can be used to measure care quality, along with examples relevant to arthroplasty, is shown [11].



measures are also useful for practical reasons in that they are usually easily collected and verifiable indicators. This observation has led to the endorsement of case-volume as a way for purchasers to identify preferred sites and improve patient outcomes [4], an approach aptly termed “follow the crowd” [2].

However well supported by evidence, many of the initial Leapfrog measures have been extraordinarily difficult to implement widely. Regionalization of services poses practical problems [1] in that it is often very hard for patients to travel long distances to seek a highest-volume center or for surgeons to provide adequate post-operative care for patients from far away. Volume benchmarks’ ability to accurately identify “best” sites has limitations [8, 9, 15, 19, 24]. Implementation of computerized order entry is a resource- and time-intensive process that may take years. Moreover, some volume standards are increasingly hard to meet because of secular changes in healthcare practice. Cardiac surgery volume benchmarks, for example, are difficult to meet as a result of improved percutaneous coronary interventions such as stenting [20].

Although we are focusing on the structural elements of the Leapfrog model, it also includes a number of quality and process measures derived in part from those recommended from the National Quality Alliance and CMS Core Measures. Adherence to Leapfrog measures is more likely at larger hospitals and those participating in other quality initiatives [11]. When adherence to Leapfrog measures is maximal, mortality is lower in acute myocardial infarction and in vascular surgery [6, 12].

Process Measures

The measurement and feedback of hospital (or surgeon) performance on specific process measures is an alternative approach to improving care. A key principle of process measurement, whether as part of guidelines from professional societies or national reporting bodies [21], is that they focus on care practices that should be followed regardless of operative volume, site of care, or surgeon. This aspect of processes means measurement of processes also means that “optimal” patient groups for each process must be defined clearly. Within “optimal” patient groups, care processes occur commonly, a feature that overcomes the statistical shortcomings of focusing on rare events such as mortality, potentially providing the ability to better detect sites with poorer performance [3]. Finally, care processes often represent clear elements of clinical practice (such as administering a medication within a certain time period) that are easily recognized as elements of everyday work and which readily form teachable skills.

There are two notable examples of process-driven quality initiatives: the Surgical Care Improvement Program (SCIP) [23] and the Physicians Quality Reporting Initiative (PQRI, www.cms.hhs.gov/pqri). SCIP arose from the Surgical Infection Prevention program as a voluntary collaborative and transitioned to a mandatory publicly reported system in 2003. The incentive for participation in SCIP was a Medicare payment withheld for nonparticipation that could be substantial depending on the size of the facility. For patients undergoing joint arthroplasty, SCIP measures represent a narrow but highly important set of complications (surgical infection prevention and deep vein thrombosis) with hospital-level performance measures currently posted on www.hospitalcompare.org.

PQRI is a recent entry into the field of process measurement. PQRI was begun as part of the 2006 Tax Relief and Health Care Act (PL 109-432), which required the establishment of a physician quality reporting system. In PQRI, physicians who report quality-measures data on claims for services furnished to Medicare beneficiaries January 1 through December 31, 2008 earn a single consolidated incentive payment of 1.5% of charges for covered Physician Fee Schedule services sometime in 2009. The PQRI specifies more than 100 potential measures of quality, the majority of which represent care processes, but a few of which (such as adoption of e-prescribing) are also structural measures. PQRI incorporates the SCIP measures as well as other measures potentially relevant to joint arthroplasty such as treatment of osteoporosis after a fracture, adequately addressing pain, and development of a care plan in conjunction with the patient. PQRI differs from SCIP in its expanded list of measures, a predominant weighting of these measures toward outpatient/clinic care, and the focus on individual physicians as the targets of the feedback and incentives.

To date, there are few data to suggest that improvement in process measure performance is associated with any improvement in patient outcomes. Publicly reported process measures for medical conditions such as acute myocardial infarction, pneumonia, or congestive heart failure have a modest association with reduced mortality but explain only a small amount of mortality variation seen across sites [5]. PQRI measures have not yet been studied, but adherence to SCIP measures does not appear to be associated with improved outcomes in general surgery; no studies have examined the impact of SCIP measures in orthopaedic surgery.

Process measures are increasingly also being seen as flawed for other reasons. First, pay-for-performance focusing on processes appears to have a weak marginal effect on improvements, particularly when most centers are improving care [16]. Second, as adherence rates rise, they will “ceiling” at 100%, making it difficult to discern

high- from low-performing sites. Third, process measures themselves may require risk adjustment to account for subtle differences in patients, even those defined as “optimal candidates” [17]. Fourth, financial incentives based on care processes tend to have weak marginal effects [16] and may magnify care disparities by taking funds away from safety net hospitals and those with higher proportions of less advantaged populations [13]. Finally, measuring individual quality measures may not be a stringent enough measure of care reliability (that is, are all care processes delivered to all patients who need them?). This shortcoming of individual process measures has led some to suggest that process measurement gives credit only if all measures are met—“all or none” measurement [18].

Measuring Outcomes as a Way to Compel Care Improvements

Outcomes are the truest end result of our care as physicians and can fall into a number of categories: positive and negative clinical outcomes, functional status, resource utilization, satisfaction with care, and health status are general domains of patient outcome.

There are a number of examples of outcome-driven healthcare improvement initiatives. Networks such as the Vermont Oxford Neonatal ICU network [22] and Project IMPAACT (an ICU collaborative) [7] are notable examples from nonsurgical specialties.

In surgery, the Veterans’ Affairs (VA) (and now private sector) National Surgery Quality Improvement Project (NSQIP) provides a notable model for orthopaedics. The NSQIP began in the VA in the mid-1990s at the behest of Congress to address higher surgical mortality in VA hospitals [14]. The VA NSQIP developed an active program of data collection (initially through paper and later through electronic sources) of both risk adjustment and outcome data, which were then used to develop risk adjustment models and benchmarks for participating VAs. Outlier sites (those in the lowest 20% of performance) underwent an audit at a distance; those in the worst 10% of performance had a site visit from NSQIP leadership. The VA NSQIP was viewed as highly successful with mortality rates falling from slightly higher than 3% to lower than 1% in non-cardiac surgery between 1995 and 2005 [14]. At the end of their first decade, NSQIP investigators found that high-performing sites shared a number of characteristics, including focus on standardization, adherence to guidelines, and focus on an interdisciplinary approach.

Focus on outcomes requires that adequate risk adjustment can be performed; often this requires collection of data not available in discharge abstract files. In addition,

outcomes often have substantial power limitations; in one study, fewer than 60% of hospitals performed enough coronary artery bypass surgery in 2 years to provide adequate sample size [3]. However, outcomes have high face validity and can be tailored to address a clinical practice specifically. Often, these outcomes can be chosen so that sample size issues can be overcome. For example, comparisons of functional status in all patients undergoing arthroplasty would have fewer power limitations than comparing mortality or need for reoperation.

Discussion

The structure-process-outcome measurement framework remains a valid starting place for defining areas where care could be improved, and there are few people who question its general usefulness in developing a measurement strategy. However, it is clear that stakeholders are increasingly focusing on initiatives which measure outcomes as the primary goal and use outcomes to define structures and processes which might be changed, rather than compelling changes in structures or quality measures only. This shift has been taking place as increasing amounts of evidence have accumulated to suggest that a multipronged approach is necessary to improve care through measurement; improved measurement will be important for surgeons and payors alike so that clear distinctions between preferred and non-preferred providers can be made.

It is fairly safe to say that the quality measurement field is not at the beginning of the end, but at the end of the beginning of its development as a science and management tool. Increasingly, effective healthcare quality and safety monitoring systems seek to achieve several goals simultaneously: to coordinate effective audit and feedback, deliver education, reengineer systems, and align goals at the patient, surgeon, hospital, and payor levels. The ability to achieve this programmatic goal will be facilitated by improvements in electronic data systems, which will both increase the availability of clinically important information (needed for risk adjustment and defining outcomes) as well as making data available without the need for costly manual chart abstraction.

In arthroplasty, there are few randomized-trial based quality or structural measures that could be applied or recommended widely; the few that exist target a narrow spectrum of problems (eg, venous thromboembolism prophylaxis). Lack of a wealth of ‘gold standard’ evidence describing all aspects of care is not unique to arthroplasty, and efforts to improve outcomes in orthopaedics will need to take the more general approach of building infrastructure necessary to both collect and compare structural, process, and outcome data effectively.

Developing a comprehensive structure-process-outcomes comparative system in arthroplasty will require investments in broad based benchmarking (or registry) projects. Given the long time frame involved in many of the outcomes of orthopaedic surgery (eg, time to redo-arthroplasty, or proportion of prosthetic devices which fail in 10 years), these efforts will need to find partners which allow patients to be tracked over time, as well as across settings of care. In addition, these systems will need to include data not commonly collected in administrative data systems, such as functional status, pain scores, or frequency of return to work. Next, a robust outcomes-based registry would be markedly enhanced by collection of data regarding the systems in place at each hospital. Finally, and most critically, standard and complete reporting of the implants used will be critical to understanding how and whether care has been improved. While this is obviously a controversial point, not knowing a potentially key 'active ingredient' to arthroplasty outcomes seems a glaring deficiency; it is doubtful that we would accept a study comparing outcomes of a number of preventative agents for thromboembolism without knowing what the drugs were, yet a similar scenario is playing out in arthroplasty.

Even the best registry will not change physician behavior unless it is linked to strong leadership from key stakeholders, beginning with the patient. This will mean that attention is paid to the effective dissemination of effective care practices when they are discovered through the usual activities of a registry (eg, benchmarking and outcomes comparison). At least as importantly, there will need to be attention paid to ineffective or harmful treatments or procedures. A key side effect of the VA portion of the National Surgical Quality Improvement Program was that care was localized away from poor-performing sites. While it is hard to see how sites could be closed or repurposed outside of the VA system, efforts which seek to remediate poor performers through site visits and efforts from professional societies may be useful.

At least as importantly, leadership in this context does not imply that leaders somehow become more effective at getting more pay for services, arguing for wider latitude for reimbursement or more flexibility in use of relatively unproven technologies. Rather, leadership in the context of improving quality and safety will first require a focus on developing generalizable and widely applicable evidence for the value of defined sets of procedures and implants, which itself requires better (or any) evidence for the marginal effectiveness of newer and costlier strategies. Once evidence is established, leadership in the era of public accountability will require a focus on standardizing practices to the extent possible and a relentless drive to introduce care that is better and more efficient for patients, not just for physicians. In this context, a traditional focus

on innovation shifts from the biomedical innovations to healthcare system innovations that improve care while retaining a clear patient focus. This is the challenge and opportunity for the 21st century and one that orthopaedics is very well positioned to adopt.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Birkmeyer JD. High-risk surgery—follow the crowd. *JAMA*. 2000;283:1191–1193.
2. Birkmeyer JD, Dimick JB. Potential benefits of the new Leapfrog standards: effect of process and outcomes measures. *Surgery*. 2004;135:569–575.
3. Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? *J Am Coll Surg*. 2004;198:626–632.
4. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003;349:2117–2127.
5. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, Wang Y, Normand SL, Spertus JA, Krumholz HM. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296:72–78.
6. Brooke BS, Perler BA, Dominici F, Makary MA, Pronovost PJ. Reduction of in-hospital mortality among California hospitals meeting Leapfrog evidence-based standards for abdominal aortic aneurysm repair. *J Vasc Surg*. 2008;47:1155–1156; discussion 1163–1154.
7. Cerner Corporation, Project IMPAACT. Available at: <http://www.cerner.com/piccm/about.html>. Accessed March 30, 2009.
8. Christian CK, Gustafson ML, Betensky RA, Daley J, Zinner MJ. The Leapfrog volume criteria may fall short in identifying high-quality surgical centers. *Ann Surg*. 2003;238:447–455; discussion 455–447.
9. Dimick JB, Finlayson SR, Birkmeyer JD. Regional availability of high-volume hospitals for major surgery. *Health Aff (Millwood)*. 2004;Suppl Web Exclusives:VAR45-53.
10. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44:Suppl:166–206.
11. Ford EW, Short JC. The impact of health system membership on patient safety initiatives. *Health Care Manage Rev*. 2008;33:13–20.
12. Jha AK, Orav EJ, Ridgway AB, Zheng J, Epstein AM. Does the Leapfrog program help identify high-quality hospitals? *Jt Comm J Qual Patient Saf*. 2008;34:318–325.
13. Karve AM, Ou FS, Lytle BL, Peterson ED. Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients. *Am Heart J*. 2008;155:571–576.
14. Khuri SF. Safety, quality, and the National Surgical Quality Improvement Program. *Am Surg*. 2006;72:994–998; discussion 1021–1030, 1133–1048.
15. Khuri SF, Henderson WG. The case against volume as a measure of quality of surgical care. *World J Surg*. 2005;29:1222–1229.
16. Lindenaier PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007;356:486–496.

17. Mehta RH, Liang L, Karve AM, Hernandez AF, Rumsfeld JS, Fonarow GC, Peterson ED. Association of patient case-mix adjustment, hospital process performance rankings, and eligibility for financial incentives. *JAMA*. 2008;300:1897–1903.
18. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA*. 2006;295:1168–1170.
19. Peterson ED, Coombs LP, DeLong ER, Haan CK, Ferguson TB. Procedural volume as a marker of quality for CABG surgery. *JAMA*. 2004;291:195–201.
20. Ricciardi R, Virnig BA, Ogilvie JW, Jr, Dahlberg PS, Selker HP, Baxter NN. Volume-outcome relationship for coronary artery bypass grafting in an era of decreasing volume. *Arch Surg*. 2008;143:338–344; discussion 344.
21. Surgical Care Improvement Program. Available at: <http://www.medqic.org/dcs/ContentServer?cid=1122904930422&pagename=Medqic%2FContent%2FParentShellTemplate&parentName=Topic&c=MQParents>. Accessed March 29, 2009.
22. Vermont Oxford Collaborative. Available at: <http://www.vtoxford.org/>. Accessed March 29, 2009.
23. Vollmer CM, Jr, Pratt W, Vanounou T, Maithel SK, Callery MP. Quality assessment in high-acuity surgery: volume and mortality are not enough. *Arch Surg*. 2007;142:371–380.
24. Ward MM, Jaana M, Wakefield DS, Ohsfeldt RL, Schneider JE, Miller T, Lei Y. What would be the effect of referral to high-volume hospitals in a largely rural state? *J Rural Health*. 2004;20:344–354.