Procedure volume as a quality measure for total joint replacement

D.S. Meredith, J.N. Katz

Orthopaedic and Arthritis Center for Outcomes Research, Department of Orthopaedic Surgery and Division of Rheumatology, Immunology and Allergy, Brigham and Women’s Hospital (JNK) and Harvard Medical School (DSM, JNK), Boston, Massachusetts, USA.

Dennis S. Meredith, MD;
Jeffrey N. Katz, MD.

Supported in part by NIH K24 AR 02123, NIH P60 AR 47782 and Harvard Medical School.

Please address correspondence to: Jeffrey N. Katz, MD, MS, Department of Orthopaedic Surgery and Division of Rheumatology, Brigham and Women’s Hospital, 75 Francis Street, Boston, MA 02115, USA.

E-mail: jnkatz@partners.org

Received and accepted on July 24, 2007.


ABSTRACT

The association between procedure volume and post-procedure outcomes such as mortality or perioperative complications has been studied since the 1970s for a wide variety of procedures. A clear understanding of the implications of procedure volume for healthcare policy is critical because payers in both the private and public sectors have begun to develop policies that incorporate procedural volume data.

In this review, we ask if procedure volume is a good surrogate measure of the quality of care. We focus specifically on literature related to total joint replacement. We have examined procedure volume as a quality measure with respect to four criteria—validity, reliability, accessibility and capacity to enact policy. The strength of procedure volume as a quality indicator in total joint replacement (and numerous other procedures) lies in its ease of access and the strength and consistency of its association with numerous important health outcomes. These include mortality, dislocation, infection, revision rates, length of stay, medical complications and improved patient satisfaction.

Although the current outcomes of total joint replacement are generally excellent, the large number of procedures performed annually makes even small improvements important. There is little information on the reliability of procedure volume assessments over time. The major weakness of procedure volume as a quality measure lies in its limited capacity to suggest cogent policy decisions. For total joint replacement, studies have failed to identify the processes that mediate the effect of increased procedure volume on patient outcomes. Consequently, there are few options for improving procedure volume beyond regionalization. More research is needed to identify the best method for regionalization and ensure that the access to health care of vulnerable populations is not compromised.

Introduction

The association between procedure volume and post-procedure outcomes such as mortality or perioperative complications has been studied since the 1970s (1). These studies of volume effects form part of the larger body of literature focused on quality improvement in health. Given that higher provider procedure volume is associated with better outcomes, it would seem natural that volume is a good marker of quality. We examine this contention carefully in this paper, using as our framework for quality the assertion that quality is “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (2, 3).

We focus our discussion of quality here on total joint replacement. In 2004, 431,000 total knee replacements (TKR) and 226,000 total hip replacements (THR) were performed in the United States (4). The annual national volume of TKRs has increased by more than 10% per annum since the 1980’s (5). This trend will accelerate as the “baby boomer” population ages. Recent projections by Kurtz et al. suggest that the demand for TKR could grow as high as 3.5 million per annum by 2030 (6).

We suggest that a good measure of quality must meet four criteria (Tab. I). The quality measure should be: (i) valid, in that it has an association with a desired health outcome; (ii) reliable, in that repeated measurements give similar results; (3) accessible, in that the data can be easily and inexpensively obtained; and (4) actionable, in that the measure points to changes in the process or structure of care that in turn yield improved outcomes. The relationship of each of these elements to procedure volume is described below.

Competing interests: none declared.
Is procedure volume a good measure of quality? / D.S. Meredith & J.N. Katz

Table 1.

<table>
<thead>
<tr>
<th>Characteristics of a good quality measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
</tr>
<tr>
<td>Reliable</td>
</tr>
<tr>
<td>Accessible</td>
</tr>
<tr>
<td>Actionable</td>
</tr>
</tbody>
</table>

Validity
Donabedian, a pioneering theorist of quality of care research, identified three types of data that can be used to evaluate the quality of care: (i) structure, (ii) process, and (iii) outcomes (7, 8). Structural data comprise the tangible characteristics of personnel, hospitals and health systems. Typical examples include the number of operating rooms, whether a physical therapy group works on site, the level of staff training and experience, and procedure volume. Process data comprise the activities of health care: i.e., what actually happens. These include the interactions between clinicians and patients, such as the surgeon’s prescribing pre-operative antibiotics, post-operative beta-blockers, and 4 weeks of physical therapy. Outcome data refer to the ultimate output of the system, such as patient mortality, infection rates, functional status and satisfaction.

Aspects of structure or process are valuable for improving quality of care if they lead to improvements in outcome or, conversely, if the identification of poor outcomes can lead to changes in structure or in the process of care. For example, outcome data on infection are most valuable if they can be related to a structure (such as the use of laminar flow hoods in operating rooms) or process (the timely delivery of peri-operative antibiotics, or the removal of Foley catheters) that can be altered to reduce infection rates in the future. Thus, the validity of procedure volume as an indicator of quality is determined by how closely the data are associated with desired health outcomes.

Procedure volume and outcomes
The literature investigating the association of procedure volume with post-procedure outcome is extensive. Studies showing an inverse relationship between hospital and/or surgeon procedure volume and the rates of mortality, functional status or surgical complications have included a wide array of procedures including coronary artery bypass grafting (9, 10), coronary angioplasty (11-14), carotid endarterectomy (15), abdominal aortic aneurysm repair (16), cancer surgery (17), pancreatic surgery (18, 19), liver transplantation (20) and cataract surgery (21). In the orthopedic literature, studies have included primary and revision arthroplasty of the hip (22-24), knee (25-27) and shoulder (28-30), as well as spinal fusion (31, 32) and hip fracture fixation (33-35). A recent comprehensive review of this literature concluded that 77% of published reports found significant inverse associations between provider volume and mortality, functional status or complications (36). The mechanisms and even the direction of these associations are not clearly defined.

The orthopedic literature includes objective outcomes such as mortality, infection, dislocation, pulmonary embolism and revision rates, as well as patient-centered outcomes that are assessed using validated instruments such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) or the Harris Hip Score (37-40) and patient satisfaction. To convey the magnitude of these volume effects, we will highlight the results from some of the larger studies in this literature that have adequate statistical adjustment for age, gender and patient comorbidities.

Mortality
Kreder et al. reviewed claims from 8,000 primary THRs in the Comprehensive Hospital Abstract Report-
a significant decrease in infection rates for high-volume surgeons performing TSR (29).

Dislocation
Katz et al. found strong associations between both hospital and surgeon volume and dislocation at 90 days following primary and revision THR (22). The rate of dislocation after revision THR in low-volume hospitals (< 5 per year) was 9.8% compared with 4.2% in high-volume hospitals. The effect of surgeon volume was more pronounced, with low-volume surgeons (< 5 per year) having a dislocation rate of 4.2% versus 1.5% for high-volume surgeons (> 50 per year) following primary THR.

Revision
Kreder et al. found that low-volume surgeons had a nearly 3-fold increase in revision rates at 90 days post-primary THR, and low-volume hospitals had a 2.2-fold increase in rates of revision following TKR at 1 and 3 years post-operation (23, 41). Losina et al. similarly found that high-volume surgeons had lower rates of THR revision at 4 years follow-up, with the effect being most pronounced in the first 18 months post-operation (44).

Length of stay
Kreder et al. found that patients of low-volume surgeons had significantly longer lengths of stay for both primary THR and TKR (23, 41, 42). The magnitude of this effect for primary THR ranged from 0.8 days longer in the CHARS database from Washington State to 2.4 days longer in the Ontario Health Insurance Plan (OHIP) database. Similarly, the TKR length of stay for low-volume surgeons was on average 1.4 days longer than for high-volume surgeons in the OHIP database. Jain et al. also demonstrated decreased lengths of stay in patients of higher volume surgeons undergoing TSR (29).

Patient-centered outcomes
Katz et al. investigated the effect of hospital and surgeon volume on the WOMAC pain and function indices, on a self-administered form of the Harris Hip Score, and on patient satisfaction at 3 years following THR (24). They found a significant association between both higher surgeon and hospital volume and increased patient satisfaction. However, there was no significant association with the WOMAC or Harris Hip scores in their multivariate analysis. Katz et al. also investigated the WOMAC indices and patient satisfaction following TCR (25) and found significant associations between high-volume hospitals and surgeons and more favorable WOMAC function scores and patient satisfaction. Volume was not associated with the WOMAC pain scores. The proportion of patients with WOMAC functional status scores < 60 was 12% in hospitals where > 200 TKRs were performed per year versus 19% in hospitals with ≤ 10 cases per year. Similarly, the proportion of patients with WOMAC functional status scores < 60 was 10% for surgeons performing > 50 TKRs per year versus 20% for surgeons performing ≤ 12 TKRs per year.

Medical complications
Medical complications that have been investigated include pulmonary embolus, pneumonia and myocardial infarction. Katz et al. found a significant inverse trend between the pooled analysis of all of the above medical complications following TKR and higher hospital volume (43). However, there was no significant association between volume and pulmonary embolus alone following THR (22).

Summary of volume effects
Taken as a whole, the procedure volume literature for common orthopedic surgeries is strongly associated with better outcomes for a number of variables including mortality, dislocation, revision, medical complications, patient satisfaction and length of stay. A recent structured review of the orthopedic literature relating to procedure volume and patient outcomes supports this conclusion (45). Specifically, it found strong associations between higher hospital volume and lower rates of mortality and hip dislocation, as well as between surgeon volume and hip dislocation. Although the results of total joint replacement are generally excellent, even for low-volume providers, the large volume of procedures performed annually makes even small improvements highly important. For instance, if all of the approximately 6,700 Medicare patients who had a primary total hip replacement in centers with annual procedure volumes of < 10 in 1995 were referred to centers where > 100 such procedures were done annually, 40 lives would have been saved (number needed to treat = 167) (46).

A notable limitation of this literature is the lack of clinical information from claims data (47, 48). Procedure complexity can be especially important for revision surgery. The large sample sizes compensate to some extent for this deficit. Nonetheless, many of the non-significant associations in the literature may arise from the lack of statistical power to examine rare adverse events following total joint replacement. Inconsistencies in this literature can also be partially explained by the limitations imposed on the data sources by the study design, such as a possibly inadequate adjustment for patient case mix and the collection of data only for acute hospital admissions. Cut-offs for low- and high-volume vary across studies. For example, Taylor et al. did not adjust for age, gender or patient co-morbidities (31). Studies have shown that low-volume hospitals treat an older population with more co-morbidities than high-volume centers (23, 30). Studies such as those by Norton et al., Hervey et al. and Jain et al. account only for events occurring during the acute hospital admission (27, 29, 49). This introduces bias for two reasons: (i) high-volume centers have a shorter length of stay, and (ii) the majority of perioperative events occur after discharge (50).

Reliability
A second component of a good quality measure is its reliability. In this review, we use reliability to refer to the reproducibility of assessing procedure volume over time. Since the data for procedure volume are derived from claims databases, there should be little variation in repeated assessments of volume.
for the same procedure over the same time period. That is, we anticipate that procedure volume assessments are not subject to the inter-test variability inherent in tests that involve interpretation, such as radiographic or pathologic readings. However, procedure volume can change over time for both surgeons and hospitals. Fluctuations can result from changes in the underlying frequency of disease (fewer fractures, fewer fracture repairs), changes in community capacity (a new surgeon moves to town and begins operating), and changes in patient preferences for undergoing the procedure. We are not aware of any peer-reviewed literature that estimates these longitudinal fluctuations in procedure volume. Consequently, it is difficult to assess their impact on the reliability of procedure volume.

Changes caused by improved outcomes are a second potential source of fluctuations in procedure volume. This theory underlies one of the potential mechanisms by which increased procedure volumes are associated with better outcomes. The idea is that a hospital with better outcomes will be recognized by patients and referring providers and, consequently, have increased volume. Katz et al. have investigated this potential mechanism using procedure volume residuals for THR (46). A procedure volume residual is the difference between the THR volume of a hospital that can be predicted on the basis of its number of beds and its teaching status, and the actual number of THR performed. If outcomes drive volume, then the residual should be associated with mortality — that is, a hospital with a relatively high mortality rate should do fewer THRs than expected and vice versa. However, residuals explained almost none (0.04%) of the variability in mortality within the model of Katz et al.

In a separate study, Hannan et al. looked at the results of public reporting of risk-adjusted mortality figures for cardiac surgery in New York State (51). When mortality data were first made available in 1989, 9% of patients were having surgery at hospitals with risk-adjusted mortality figures significantly higher than the state average. Four years later, the number of patients having surgery at centers with a significantly higher reported mortality was 10%. Although causality cannot be established with certainty in the absence of a randomized controlled trial, these results suggest that outcomes do not drive procedure volume changes over time. In summary, further research is needed to quantify the magnitude of procedure volume changes over time, but we cannot identify any quality of care mechanisms that would directly alter procedure volume over time.

Accessibility
A major advantage of procedure volume data for assessing quality of care is that they can be accessed easily and relatively inexpensively. Currently, there are multiple sources for procedure volume data. Claims databases referenced in this review include Medicare (22, 43), the Canadian Institute for Health Information (CIHI) (42), the Ontario Health Insurance Plan (OHIP) database (41), the Statewide Planning and Research Cooperative System (SPARCS) database from the New York State Department of Health (30), the Nationwide Inpatient Sample (NIS) database from the Healthcare Cost and Utilization Project (HCUP) (29), the Maryland Health Services Cost Review Commission hospital discharge database (28), and the Comprehensive Hospital Abstract Reporting System (CHARS) database from the Washington State Department of Health (23). These resources encompass data on a statewide or nationwide scale. They can be accessed less expensively and with less administrative effort than many other sources for quality of care measures such as process data or functional outcome questionnaires.

Policy implications
A clear understanding of the implications of procedure volume for healthcare policy is critical, because payers in both the private and public sectors have started paying attention to volume. For instance, the Leapfrog Group, a consortium of major businesses dedicated to improving healthcare quality and efficiency, has recommended that its client organizations use only high-volume providers for selected surgical procedures (52). Similarly, The Centers for Medicare and Medicaid Services (CMS), which manages the Medicare program, has initiated a pilot program that Designates centers of excellence for total hip and knee replacement surgery. However, the debate on this issue is far from settled. In this section we will focus on the difficulties of implementing policy to improve outcomes based on procedure volume data.

The most frequently discussed policy intervention to address procedure volume effects is regionalization. This refers to preferential referral of procedures to selected centers to increase procedure volume. One of the oldest forms of regionalization is government certificate-of-need (CON) regulations. CON programs concentrate expensive healthcare services in a limited number of institutions by requiring prior approval before these services can be offered. CON laws for cardiac surgery are currently present in 26 states and Washington DC. A recent study by Vaughn-Sarrazin et al. showed that CON laws have a significant beneficial effect, both increasing coronary artery bypass graft (CABG) procedural volume and improving mortality for elderly Medicare patients (53). However, a subsequent study by DiSesa et al. using the Society of Thoracic Surgeons’ National Cardiac Database did not find that CON programs had any effect on CABG-related mortality after adjustment for the patient mix, region and population density (54). CON programs comprise one of the few examples where we have data on the effect of policies aimed at increasing procedure volume. More research is needed to document the effect of regionalization programs on desired health outcomes such as mortality. One limitation of structural data such as procedure volume is that they do not offer as many options to intervene as process data. It is generally presumed that procedure volume is a proxy for one or multiple processes of care that improve outcomes. For example, a study by Sollano et al. found that...
the effect of procedural volume for CABG-related mortality was reduced to an insignificant level when the authors adjusted for the effects of New York State’s CABG quality improvement program (55), suggesting that high-volume hospitals performed well because they had adopted the quality improvement program. However, in the orthopedic literature efforts to identify processes of care that explain a significant proportion of the volume effect have not yielded plausible explanations (56). Solomon et al. examined a variety of hospital characteristics including teaching program status, JCAHO accreditation, dedicated orthopedic surgery operating rooms, care pathways, nurse-to-patient ratios, physical therapy and the presence of laminar exhaust systems. The association of these factors with outcome was very modest and far less potent than the association with surgeon volume. Consequently, the only consistently supported mechanism by which procedure volumes exert their effect on outcomes of orthopedic procedures is the idea that “practice makes perfect.” This represents a limitation in that low-volume hospitals or surgeons may face barriers that curb their ability to become high-volume providers.

The foregoing suggests that volume effects cannot be remedied at the process level given current knowledge. How then can volume effects be addressed with policy remedies? Payer level interventions present the most obvious first choice. Payers [both federal (Medicare, Medicaid) and private] can steer patients to higher volume centers if they wish, and several such regionalization initiatives have been launched. However, restricting the patient’s choice of hospitals may diminish patient satisfaction regardless of the surgical outcome (57).

Another option is to maximize patient choice by informing patients and asking them to make decisions. However, survey data show that very few elderly patients considered information on volume or mortality in deciding where to have surgery, even when the data was publicly available (58, 59). This is an important limitation, indicating that the factors necessary to effect a change in health behavior are frequently complex. One of the most widely used conceptual frameworks of health behavior, the Health Belief Model, stipulates that to effect a change the patient must be convinced that: (i) they are at risk for a potentially adverse event such as surgical complications; (ii) the adverse outcomes are severe enough to merit concern; (iii) alternatives are available to reduce the risk of an adverse event; and (iv) their own personal barriers to implementing a given alternative can be overcome (60). It is unlikely that passive intervention such as mailings from the payer or posting data on a website will facilitate these processes and effect a change in behavior. Active intervention is needed at multiple levels of the healthcare system to integrate procedure volume effectively into everyday decision-making about where to have a procedure performed. This constitutes an important area for further work.

Of note, 73% of the patients surveyed said that their referring provider played a dominant or equal role in deciding where to have surgery. Consequently, the dissemination of procedure volume information to referring providers, combined with education on how to interpret this data, could have the greatest impact while maintaining patient choice.

Another issue for policy-making is whether there is a threshold level where the maximum effect from volume is seen. However, there is no definitive answer to this question at this time. The studies of Katz and colleagues suggest that there may be no discrete threshold for THR outcomes (22), but that there may indeed be a threshold (around 25 cases per year) for TKR outcomes (43).

Finally, while there may be a strong association between procedure volume and improved outcomes, there could also be unforeseen consequences to regionalization policies that need to be considered. Approximately 50% of all the hospitals providing TKR within the Medicare claims database are low-volume providers (< 26 per year), accounting for 13% of all TKRs performed annually (61). Recent studies show that the population served by these low-volume hospitals is comparatively poor, rural, less educated, and elderly with a higher prevalence of ethnic and racial minorities (62). Such “vulnerable” populations already have lower rates of total joint replacement after adjustments for age and patient co-morbidities (63-66) (see chapter by Emjuaiwe, Kwoh, et al. in this volume). Studies have shown that some patients would refuse surgery entirely if their only option was a distant or unfamiliar center (67). Alternately, a significant delay in time to surgery for total joint replacement is also associated with significantly worse outcomes (47, 48). Regionalization policies must take into account these vulnerable populations so that we do not trade large reductions in access to care for small improvements in perioperative complication rates.

Conclusion

Increasingly, payers in both the private and public sectors are using procedure volume to enhance patient safety. We have examined procedure volume as a quality measure with respect to four criteria: validity, reliability, accessibility and actionability – the capacity to enact policy based on procedure volume. A summary of our conclusions can be found in Table II. The strength of procedure volume lies in its ease of access and the strength and consistency of its association with important health outcomes, including mortality, dislocation, infection, revision rates, length of stay, medical complications, and improved patient satisfaction, for numerous procedures including total joint replacement. Although the current outcomes of total joint replacement are generally excellent, the large number of procedures performed annually renders even small advances in safety important. There is very little information on the reliability of procedure volume assessments over time. The major weaknesses of procedure volume as a quality measure arise from their limited capacity to suggest policy remedies. For total joint replacement, studies have failed to identify the mediating processes that could explain why increased volume leads to better outcomes. Consequently, there are few options for improving...
procedure volume beyond regionalization policy. More research is needed to identify the best strategies for regionalization and to ensure that the access to health care of vulnerable populations is not compromised.

References

11. HANNA EL, CAHILL GV, DELONG ER et al.: Coro-
14. KIMBEL SE, HERLIN JA, LASKEY WK: The relationship between coronary angioplasty
procedure volume and major complications. JAMA 1995; 274: 1137-42.
16. MANHEIM LM, SOHN MW, PHINGLIAS J, UIKKI M, PARKER MA, PEACH WH: Hospital vascular surgery volume and procedure mor-
17. BEGG CB, CRAMER LD, HOSKINS WJ, BRENN-
MAN MF: Impact of hospital volume on op-
19. IMPERATO PJ, NENNER RP, STARR HA, WILL-
TO, ROSENBERG CR, DEARIE MB: The ef-
20. LEWAKIN EH, KOULHERIS JP, MEHRIE MA, SCHULAK JA, HUNSICKER LG: The effect of the volume of procedures at transplantation centers on mortality after liver transplanta-
22. KATZ JN, LUNDIN E, HARRELL J et al.: As-
23. KREDER HR, DERO RA, KOEPSSELL T, SWJNTEJKOWSKI MP, KLEINER W: Relation-
24. KATZ JN, PHILLIPS CB, BARON JA et al.: As-
association between hospital and surgeon volume of total hip replacement with functional status and satisfaction three years following sur-
25. KATZ JN, MAHUMED NN, HARIN JA et al.: As-
26. MULDOWNE JH, LIHENIK LJ, LUNDIN E et al.: Development of a preliminary index that pre-
dicts adverse events after total knee replace-
27. NORTON EC, GARKINDEL SA, MCQUAY LJ et al.: The effect of hospital volume on the in-hospital complication rate in knee replace-
28. HAMMOND JW, QUEALE WS, KIM TK, MC-
FARLAND EA: Surgeon experience and cli-
nical and economic outcomes for shoulder ar-

Table II.

Summary of procedure volume as a quality measure for total joint replacement

<table>
<thead>
<tr>
<th>Valid</th>
<th>In larger studies with adequate statistical adjustment for patient age and co-morbidities, procedure volume at the surgeon and hospital levels has been shown to have a significant association with multiple important health outcomes after total joint replacement, including:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mortality</td>
<td></td>
</tr>
<tr>
<td>• Dislocation</td>
<td></td>
</tr>
<tr>
<td>• Infection requiring debridement or implant extraction</td>
<td></td>
</tr>
<tr>
<td>• Medical complications (pneumonia, myocardial infarction and pulmonary embolism)</td>
<td></td>
</tr>
<tr>
<td>• Revision rates</td>
<td></td>
</tr>
<tr>
<td>• Length of stay</td>
<td></td>
</tr>
<tr>
<td>• Functional outcomes (WOMAC scores for TKR only)</td>
<td></td>
</tr>
<tr>
<td>• Patient satisfaction</td>
<td></td>
</tr>
</tbody>
</table>

Reliable: The variation in hospital volume from year to year is unclear. However, outcomes do not appear to directly influence volume.

Accessible: Procedure volume data are easily and inexpensively accessed from multiple state and nation-wide databases.

Actionable: • No processes have been identified that can explain the beneficial effect of increased procedure volume.
• Regionalization is the most commonly discussed method for increasing procedure volume. One of the few regionalization policies for which data are available is the so-called Certificate of Need (CUN) Regulations. Regionalization has had inconclusive results with respect to improved health outcomes for cardiac surgery.
• Focusing on referring physicians is likely to be the most effective approach in directing patients to high-volume hospitals. Payer-level interventions that restrict patient choice may reduce patient dissatisfaction.
• Vulnerable populations such as the poor, uneducated, elderly and ethnic/racial minorities are most likely to be adversely affected by regionalization policies. Their needs should be taken into consideration.

Is procedure volume a good measure of quality? / D.S. Meredith & J.N. Katz


