Annals of Internal Medicine

Editorial

Quality Wrapped in Volume Inside a Hospital

Studies that documented surgical outcome differences proliferated in the late 1990s. High-volume hospitals (that is, those that performed larger numbers of a particular procedure) tended to have better outcomes than their lower-volume peers (1). Nested within these findings was a similar volume–outcome relationship for surgeons (2). When low- and high-volume hospitals were compared, some outcome gaps were very large. Begg and colleagues (3) reported a 14% absolute difference in 30-day mortality rates among patients who had esophagectomy for esophageal cancer. Other gaps were modest. Schrag and coworkers (4) reported a 2% absolute difference in mortality rates 30 days after colorectal cancer surgery. Investigators have also identified a volume–outcome relationship for coronary artery bypass grafting, which has been weak in some studies and stronger in others (5).

Policymakers noticed this volume–outcome relationship. A seminal Institute of Medicine report recommended that "patients undergoing procedures that are technically difficult to perform and have been associated with higher mortality rates in lower-volume settings receive care at facilities with extensive experience (i.e., high-volume facilities)" (6). At about the same time, a consortium of large, self-insured employers began the not-for-profit Leapfrog Group (www.leapfroggroup.org), which encouraged consumers to seek care in institutions that met certain quality standards, including annual procedure volume for some high-risk surgeries (7).

The prevailing explanation for the superior outcomes at high-volume centers is that volume is a proxy for experience, which in turn enhances outcomes. The alternative explanation—patients flock to centers that have superior outcomes, which increases those centers' volumes—is not plausible because hospitals do not routinely report their outcomes. Because experience seems to matter, investigators have sought the key structural and process components that emerge through experience (8).

In this issue, Auerbach and colleagues (9) purport to explain most of the volume–outcome relationships in coronary artery bypass grafting. The authors analyzed the outcomes of more than 80 000 patients who had bypass surgery at 1 of 164 hospitals. Receiving appropriate care, as measured by 6 evidence-based quality measures, seemed to explain the relation between volume and outcome. This finding makes the article an important contribution to the literature on both quality of care and the volume–outcome relationship.

On the quality front, the study is another grim reminder of how often hospitals do not take the most basic steps to enhance patient outcomes after a frequently performed major procedure, even when the hospitals have dedicated themselves to quality improvement (the study hospitals were members of Premier [Charlotte, North Carolina], an organization that helps hospitals improve quality of care). Many patients in the study did not receive basic, inexpensive, long-established, and evidence-based risk-reducing treatments and interventions. A worrisome 20% did not receive a β-blocker within 2 days after surgery. Hospitals did even worse on most of the other quality measures. One take-home message is that the care provided in U.S. hospitals (for even common conditions) continues to be seriously deficient.

On the volume front, the authors conclude that adherence to quality measures largely explains the link between volume and outcome. The conclusion suggests that quality improvement efforts should improve outcomes after coronary artery bypass grafting, as it has in other care settings (10). Although this conclusion is viscerally appealing, the study was not an ideal test of the authors' hypothesis.

First, the relationship between volume and outcome in the authors' data set is very weak, which limits the study's power to detect a residual volume–outcome relationship after controlling for quality (that is, the risk for type II error is high). The lack of a meaningful volume–outcome relation in their study may simply be an anomaly of their data set. Alternatively, adoption of routine processes may be driving cardiac surgical care toward equivalence across care settings, regardless of hospital caseload. Indirect evidence supports this possibility: The mortality rate reported here, which represents operations done from 2003 to 2005, is considerably less (1.7%) than what Peterson and coworkers' (5) reported in a similar sample and range of hospital volumes observed from 2000 to 2001 (2.7%).

Second, the choice of predictors may exaggerate the apparent predictive superiority of quality over volume. Although the authors measured volume at the hospital (or surgeon) level, they analyzed quality of care at the patient level. Because quality varies between patients within a hospital, patient-level quality measures have more power to predict outcome than the average quality of care within the hospital, which would have been the more appropriate comparator. A familiar analogy is the link between socioeconomic status and health outcomes. Studies have shown that community-level measures of socioeconomic status, which are summary measures of the socioeconomic status of the individuals within the community, are weaker predictors of health status than the individual measures themselves (11).

Finally, the study compares 2 predictors—quality of care and volume—whose individual-level values range across dissimilar scales: quality points and case counts. When the authors conclude that quality is more predictive than volume, this conclusion has meaning only if the authors can statistically evaluate the effect on outcome when each respective predictor changes by the same amount. Yet, comparing a 1-unit change in quality to a 1-unit change in volume has no real meaning, in that the units of the 2 measures in this analysis are constructed by empirically dividing the distribution of each into its own quartile

© 2009 American College of Physicians

groups. Put another way, a 1-unit change of either predictor equates to a move from 1 quartile to an adjacent quartile in the study, but the range of values of each predictor both within and across the quartiles is a feature of the values of each that happened to appear in the data set—not something intrinsic to the actual predictors' importance.

Consider an alternative scenario in which the relationship between quality and outcome is unaltered, but quality itself was more consistent across patients and hospitals than in the authors' data set. Under this scenario, the average quality across the 4 quartiles would be more similar, and therefore outcome would seem to vary less between the quality quartiles. At the limit, if quality were exactly the same across all patients, then quality would seem to have 0 discriminatory power because variation in the predictor underlies the ability to detect a correlation between it and the outcome. Conversely, if the range of volumes in the study had been wider, with, for instance, the lowest and highest quartiles containing only hospitals with extremely low and high volumes, respectively, then volume would seem to be a much stronger predictor of patient outcome.

Consequently, a different range of values of the predictors within the data set could cause the results to look different or reversed, even when the underlying predictive power of a certain level of quality or a certain annual volume was unchanged. In the context of Auerbach and colleagues' study, the authors' findings reflect an amalgam of the "true" relative importance of volume and quality and the happenstance categories in which the predictors were placed. To isolate the "true" from the "happenstance" would require placing the 2 measures on the same scale. To do so is difficult, but it is important to confront when the objective of the analysis is to improve patient outcomes, as it is in this case.

Even with these limitations, the study contains several insightful findings. First, despite the statistical concerns I raise about the comparison between volume and quality, the implication that quality is more strongly associated with outcome than volume, is important. It suggests that regionalization of cardiac surgery may not enhance outcomes to the same degree that quality improvement efforts could. However, to be cautious, we should not generalize this conclusion beyond coronary artery bypass grafting done in hospitals that are similar to the ones in the authors' data set. On the other hand, the data on quality of care are persuasive and generalizable beyond the study sample. Studies in many hospitals have shown that patients who have coronary artery bypass grafting have better outcomes when they receive specific care processes and worse outcomes when they do not. Moreover, quality is too often unacceptably low: Hospitals regularly do not provide established evidence-based interventions that improve outcome.

Policymakers should note these latter findings. Payers could penalize hospitals that do not provide the services that they should when performing coronary artery bypass grafting. Such an initiative could build by analogy on Medicare's recent decision not to pay for avoidable complications, based on the theory that the complication should have never occurred (12). If hospitals had to pay penalties (or receive lower reimbursement) when they did not deliver well-established, outcome-enhancing services, payers could justify the penalty by saying that failure to deliver a beneficial, safe service should never occur. An alternative is to try to improve post–coronary artery bypass grafting outcomes by telling patients that they should demand the 6 services the authors analyzed. Physicians could tell their patients to go to the hospital accompanied by their loved one before coronary artery bypass grafting. That individual, checklist in hand, would advocate for the patient. Health care providers would only grudgingly accept either of these solutions; however, our feelings are a minor concern when set alongside the patients who are harmed daily through our system's shortcomings.

Peter B. Bach, MD, MAPP
Memorial Sloan-Kettering Cancer Center
New York, NY 10065

Potential Financial Conflicts of Interest: *Other:* Member of the Committee on Performance Measurement of the National Committee for Quality Assurance.

Requests for Single Reprints: Peter B. Bach, MD, MAPP, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 221, New York, NY 10065.

Ann Intern Med. 2009;150:729-730.

References

1. **Halm EA, Lee C, Chassin MR.** Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med. 2002;137:511-20. [PMID: 12230353]
2. **Birkmeyer JD, Siewers AE, Marth NJ, Goodman DC.** Regionalization of high-risk surgery and implications for patient travel times. JAMA. 2003;290: 2703-8. [PMID: 14645312]
3. **Begg CB, Cramer LD, Hoskins WJ, Brennan MF.** Impact of hospital volume on operative mortality for major cancer surgery. JAMA. 1998;280:1747-51. [PMID: 9842949]
4. **Schrag D, Cramer LD, Bach PB, Cohen AM, Warren JL, Begg CB.** Influence of hospital procedure volume on outcomes following surgery for colon cancer. JAMA. 2000;284:3028-35. [PMID: 11122590]
5. **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. [PMID: 14722145]
6. **Hewitt M.** Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality. Washington, DC: National Academy Pr; 2000.
7. **Birkmeyer JD, Finlayson EV, Birkmeyer CM.** Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. Surgery. 2001; 130:415-22. [PMID: 11562662]
8. **Bach PB, Cramer LD, Schrag D, Downey RJ, Gelfand SE, Begg CB.** The influence of hospital volume on survival after resection for lung cancer. N Engl J Med. 2001;345:181-8. [PMID: 11463014]
9. **Auerbach AD, Hilton JF, Maselli J, Pekow PS, Rothberg MB, Lindenauer PK.** Shop for quality or volume? Volume, quality, and outcomes of coronary artery bypass surgery. Ann Intern Med. 2009;150:696-704.
10. **Sehgal AR.** Impact of quality improvement efforts on race and sex disparities in hemodialysis. JAMA. 2003;289:996-1000. [PMID: 12597751]
11. **Bach PB, Guadagnoli E, Schrag D, Schussler N, Warren JL.** Patient demographic and socioeconomic characteristics in the seer-medicare database applications and limitations. MedCare. 2002;40(8 Suppl):19-25. [PMID: 12187164]
12. **Rosenthal MB.** Nonpayment for performance? Medicare's new reimbursement rule. N Engl J Med. 2007;357:1573-5. [PMID: 17942869]