The Volume–Outcome Relationship: Don't Believe Everything You See

Caprice K. Christian, M.D., M.P.H.,[1] Michael L. Gustafson, M.D., M.B.A.,[1] Rebecca A. Betensky, Ph.D,[2]
Jennifer Daley, M.D.,[3,4] Michael J. Zinner, M.D.[1]

[1]Department of Surgery, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA
[2]Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave., Boston, Massachusetts 02115, USA
[3]Tenet Healthcare System, 13737 Noel Rd., Suite 100, Dallas, Texas 75240, USA
[4]Department of Community and Family Medicine, Dartmouth Medical School, Hinman Box 7250, Strasenburg Hall, Hanover,
New Hampshire 03755, USA

Abstract. This paper investigates methodological limitations of the volume–outcome relationship. A brief overview of quality measurement is followed by a discussion of two important aspects of the relationship.

As the literature documenting a consistent relationship between procedural volume and outcome continues to grow, there are advocates of moving forward to volume-based policy changes [1–3]. For example, the Leapfrog Group is a consortium of healthcare purchasers who support selective referral to high-volume institutions [4]. To emphasize the advantages of selective referral, advocates point to the number of lives that would be saved if all patients were treated at high-volume institutions [3, 5–7]. There are also opponents to policy initiatives based on the volume–outcome relationship [8–11]. These focus primarily on the implications of such policy changes, including long travel times for patients in rural areas, the creation of a two-tiered medical system for those rural patients unable or unwilling to travel, unintended alterations of referral patterns, a lack of continuity in postoperative care, and the possibility of further overwhelming already busy high-volume centers. It is possible that the high-volume centers in certain geographic locations would be unable to handle the increase in demand.

Yet there are other important reasons not to move forward with policy changes based solely on the volume–mortality relationship, reasons related to the relationship itself. First, the etiology of the relationship between volume and outcome is still not well understood [9, 11]. The idea of "practice-makes-perfect" has obvious face validity, but studies do not support it over alternate theories [12]. Another explanation is based on "selective-referral patterns": surgeons and institutions with better outcomes receive more referrals, leading to higher volumes. Additionally, it is widely recognized that volume is not a direct measure of quality.

Correspondence to: Michael J. Zinner, M.D., e-mail: mzinner@part-ners.org

Rather, it is a proxy for other measures, such as structure and process characteristics, which more accurately reflect quality of care.

In this article we investigate the methodological limitations relating to the volume–outcome relationship. We begin with an overview of the strengths and limitations of various quality measures. In particular, we describe why mortality alone is a limited measure. We then look at some of the statistical limitations in analyses of the relationship between surgical volume and outcome. Our goal is to highlight some of the inherent difficulties in the volume–outcome relationship as reasons to be wary of making policy changes at this point.

Is Mortality the Gold Standard for Quality Measurement?

There are three accepted domains of quality of care: structure, process, and outcome [13, 14]. Even among these accepted aspects of quality, debate exists about which is the best reflection of quality [15–17]. Outcomes are easier to measure and compare; yet process measures directly reflect whether the appropriate care is given at the appropriate time. Outcome measures are subject to many other contributing factors, even when risk-adjusted, and process measures are more difficult to measure and define, especially in surgery, where little work has been done in this area.

In surgery, many suggested quality measures relate to structure. The teaching status of hospitals, the existence of specialized intensive care units and operating rooms, and staffing ratios are all examples of structural characteristics that could contribute to the observed volume–outcome relationship, but studies to support this are limited [6, 18]. A recent artcicle investigates which hospital characteristics, including house staff and nursing staff ratios, teaching status, geographic location, and hospital ownership, contribute to the volume–outcome relationship [19]. Importantly, the authors found that for procedures with equivalent staffing, outcomes appeared to be equivalent between high- and low-volume institutions for high-risk procedures.

1242

World J. Surg. Vol. 29, No. 10, October 2005

The value of process indicators in surgical quality measurement is also gaining interest. Much of the earlier work relates to clinical pathways and cardiovascular procedures [20–24]. There are a few studies that have used process measurement as a means of quality improvement, primarily by studying "high quality" outliers [19, 25, 26]. The Leapfrog Group, the consortium that supports selective referral to high-volume hospitals based on designated volume thresholds, has recently begun to investigate the use of process measures [4]. There are a number of reasons why this is difficult. First, as mentioned above, there is not much data available to define these measures. Second, it will take years for hospitals to develop and institute the necessary infrastructure for such process measurement. Finally, once these quality indicators have been identified and measured, the national resources may not be sufficient. For example, the Leapfrog Group is calling for intensivist staffing of all hospitals, yet there are not enough intensivists being trained to fill those positions. In the meantime, surgical quality measurement appears to fall to outcomes.

Mortality is the most attractive outcome measure due to its ease of measurement, particularly in administrative databases. Yet, mortality is subject to the limitations of any outcome measure, and additionally to the limitation that it is a rare event for most procedures in modern surgical practice, leading to complexities with analysis [27]. Given the lack of consensus on using outcome as opposed to process measures and the limitations of mortality as a quality measure, it seems dangerous to take this one step further to use a proxy such as volume to estimate quality based solely on its relationship with the outcome of mortality.

Level of Analysis

Volume at the Level of the Surgeon or the Institution

To address the limitations of the volume–outcome analysis, we begin with the level of analysis. First, it is necessary to determine which is more important, the individual surgeon or the institution. The literature on the relationship between individual provider volume and outcome is not as consistent as the data supporting the institutional volume–outcome relationship [28–32]. Surgeon volume and institutional volume represent very different aspects of quality of care. On the one hand, surgeon volume is a proxy for such individual human factors as technical skill and quality of decision making. Hospital volume, on the other hand, reflects institutional characteristics, as previously mentioned.

Outcome at the Level of the Patient versus the Institution

One can look at the relationship between the institutional volume and the institutional mortality rate or mortality at an individual patient level. Mounting evidence suggests that the most valid statistical approach requires analysis at the level of the individual patient. In other words, one must investigate the effect of institutional volume on mortality using regression analysis, with the patient as the unit of analysis. However, in practicality, one must also understand what is occurring at an institutional level, as the goal is to correlate the performance of the hospital or the quality of care delivered by that institution with case volume. The nature of this relationship is not yet understood, and one could hypothesize that it may take on any number of forms. It is unclear at this point whether the relationship between volume and mor-
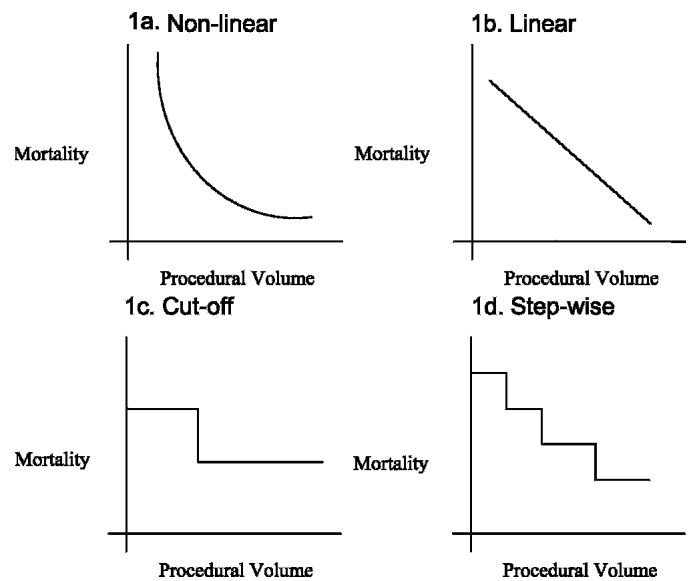

Fig. 1. Possible relationships between volume and outcome.

tality is continuous, step-wise, or has a single clear cut-off (see Fig. 1). The idea of selective referral, as proposed by the Leapfrog Group, assumes a single cut-off (Fig. 1C), yet there is not yet evidence to support this relationship over the others.

Most studies on volume–outcome divide the patients into groups of equal volume (for example, top 25%, upper-middle 25%, lower-middle 25%, lower 25%) and compare mortality between these groups. This is a well-accepted analytic approach to increasing the power of a study by increasing the number of cases (N) in each group; however, there are several problems with this approach. First, if broken into equal groups based on institution, there is large variability between the number of patients reflected in a group; however if broken into equal number of patients, there is substantial variability in the volume range per group. Often, the results of these analyses are reported as the difference in mortality across the extreme quartiles or quintiles. Despite this, some policy initiatives are advocating the use of a single strict volume cut-off to discriminate between "high quality" and "low quality" (as reflected by mortality rate). Studies to date have not addressed the existence of a single volume threshold. The idea of a single "discriminator" is appealing, but it may be overly simplistic. A recent study by the authors suggests that there may be an identifiable single cut-off; however, further work is needed to see if these thresholds are widely applicable [33].

The Variability Issue

A basic statistical principle of a Bernoulli event (e.g., coin flip) is that as the number of observations (N) increases, the variability of the estimate of the rate of the event decreases and the estimate approaches the true rate. In the example of a fair coin flip, the true rate of the flip resulting in heads is 50%. Yet, if the coin is flipped only a few times, the observed rate of heads will likely vary greatly from this true rate. Only when N is sufficiently large, can we be guaranteed that the observed rate reflects the true rate. The definition of what is meant by "sufficiently large" is a function of the underlying true rate of occurrence. The rarer an occurrence,
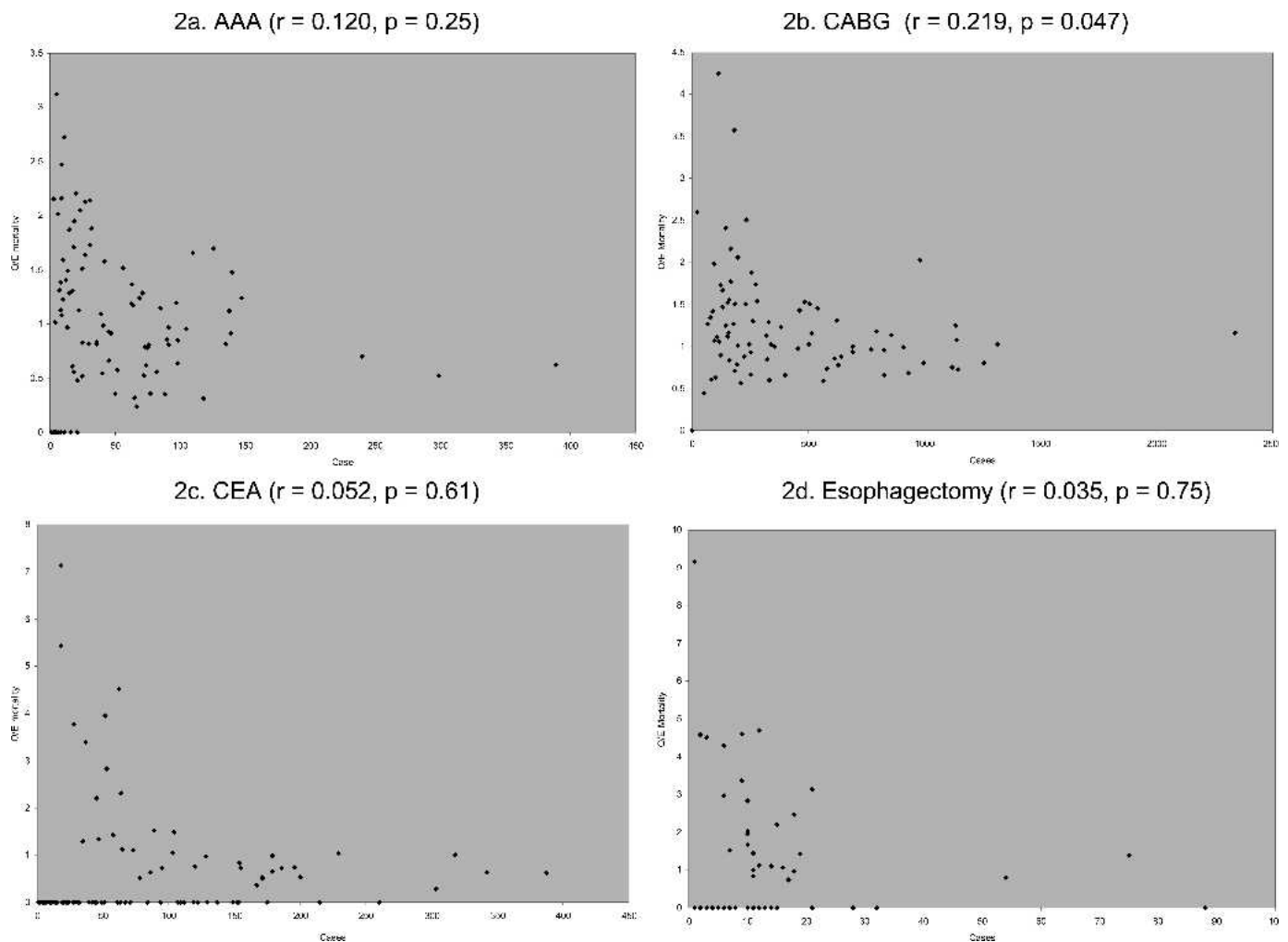
Fig. 2. Scatter plot of O/E mortality by procedural volume. The ratio of the observed raw mortality to the expected mortality rate (based on patient-specific characteristics) is plotted against procedural volume for each of the four procedures. The Pearson correlation coefficient (r) and p-value are included in parentheses for each procedure. AAA: abdominal aortic aneurysm repair; CABG: coronary artery bypass grafting; CEA: carotid endarterectomy

the greater N will need to be to assure the observed rate reflects the true underlying rate.

This principle can be applied to mortality rates to show the difficulty in analyzing the relationship between volume and mortality. In low-volume institutions, where only a few cases are done, there is a high degree of variability in the observed mortality rate that may not truly reflect the quality of care. If a given hospital performs one procedure, there are only two possibilities for the observed mortality rate: 0% (the patient does not die) or 100% (the patient dies). As another example, let us assume that a given hospital performs 5 procedures annually and the true underlying mortality rate is 0.7%. The only possible values for the observed mortality at this hospital are 0 (0 of 5 patients die), 20% (1 of 5 patients die), 40% (2 of 5 patients die), 60% (3 of 5 patients die), 80% (4 of 5 patients die), and 100% (all patients die), with respective probabilities of 0.73, 0.23, 0.03, 6×10^{-5}, 8×10^{-7}. Thus, 27% of the time, the measured mortality rate for this hospital is 20% or greater despite the fact that the true mortality rate was set at 0.7% in this hypothetical example. In contrast, if the hospital now performs 10 procedures in a given year and the true mortality

rate is stable at 0.7%, only 12% of the time will the measured mortality rate be 20% or greater. If the annual volume is 15, the observed mortality rate will be greater than 20% only 6% of the time. When the number of cases performed at a hospital is low, the most common mortality rate will be zero and a few outliers will drive the overall relationship. If a death does occur, the observed mortality is going to vary substantially from the true underlying mortality rate. As the volume increases, the observed mortality rate will naturally converge toward the true mortality rate. The true mortality rate can be estimated by the expected mortality rate. This suggests that crude volume–outcomes analyses may be biased against smaller hospitals.

Figure 2, which plots adjusted mortality against volume, illustrates these points. Adjusted mortality is a ratio of the observed mortality rate to the expected mortality rate (which is an estimate of the true mortality rate). The observed to expected mortality ratios and volumes in Figure 2 are drawn from an analysis of the University HealthSystem Consortium (UHC) Clinical Database for the year 2000 [34]. The UHC computes an expected mortality rate based on the sum of the probability of death for the patients

1244

World J. Surg. Vol. 29, No. 10, October 2005

treated at a given institution. The individual patient's probability of death is calculated using logistic regression modeling of that patient's preoperative risk factors, such as comorbidities and demographics. The probabilities are then added together to calculate the expected mortality rate for a given institution. To the eye, these data seem to demonstrate a highly correlated inverse relationship, but the correlation coefficient is not significant for three of the four graphs. Only for coronary artery bypass grafting (CABG) is there a weak ($r = -0.219$) but statistically significant inverse relationship ($p = 0.047$). This is in part due to the large number of institutions with zero mortality at the lower end of the volume spectrum. Progressing along the abscissa, the variability between institutions decreases and the graph converges.

Conclusions

Important policy changes, such as selective referral, are being suggested based on the relationship between volume and outcome. Volume offers an easy, simple measure that can be easily understood and employed by both healthcare providers and consumers. However, the use of volume as a quality indicator is not as straightforward as it may seem. Mortality is itself a limited measure of quality, and the methodologies used to evaluate the volume–outcome relationship may not be as simple as they at first seem. There is certainly value in continuing to investigate the volume–outcome relationship, however. Studies that lead to better understanding of the underlying institutional factors that lead to the difference in outcomes are crucial. In the meantime, it seems premature to be moving forward with referrals based on volume alone when there are still many issues to be resolved.

We strongly support the profession, the consumers, and the purchasers of healthcare for their desire to truly understand high quality health care, but we also caution against oversimplified solutions that do not accurately address those concerns. We, as a profession, should continually strive to improve the delivery of care to all of our patients, and understanding quality is an important aspect of that commitment.

References

1. Epstein AM. Volume and outcome—it is time to move ahead. N. Engl. J. Med. 2002;346:1161–1164
2. Halm EA, Lee C, Chassin M. How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature. Washington, DC: National Institute of Medicine, 2000
3. Dudley RA, Johansen KL, Brand R, et al. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. J. A. M. A. 2000;283:1159–1166
4. http://www.leapfroggroup.org
5. Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. Surgery 2001;130:415–422
6. Birkmeyer JD. Should we regionalize major surgery? Potential benefits and policy considerations. J. Am. Coll. Surg. 2000;190:341–349
7. Gordon TA, Bowman HM, Tielsch JM, et al. Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. Ann. Surg. 1998;228:71–78
8. Russell TR. Invited commentary: Volume standards for high-risk operations: an American College of Surgeons' view. Surgery 2001; 130:423–424
9. Dudley RA, Johansen KL. Invited commentary: physician responses to purchaser quality initiatives for surgical procedures. Surgery 2001;130:425–428
10. Khuri SF. Invited commentary: Surgeons, not General Motors, should set standards for surgical care. Surgery 2001;130:429–431
11. Daley J. Invited commentary: Quality of care and the volume–outcome relationship—what's next for surgery? Surgery 2002;131:16–18
12. Luft HS, Hunt SS, Maerki SC. The volume–outcome relationship: practice-makes-perfect or selective referral patterns? Health Serv. Res. 1987;22:157–182
13. Donabedian A. The quality of care: how can it be assessed? J. A. M. A. 1988;1988:1743–1748
14. Donabedian A. Evaluating the quality of healthcare. Millbank Mem. Fund Q. 1966;44:S166–S206
15. Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. J. A. M. A. 2000;284:1670–1676
16. Palmer RH. Process-based measures of quality: the need for detailed clinical data in large health care databases. Ann. Intern. Med. 1997;127:733–738
17. Palmer RH. Using health outcomes data to compare plans, networks, and providers. Int. J. Quality Healthcare 1998;10:477–483
18. Khuri SF, Najjar SF, Daley J, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. Ann. Surg. 2001;234:370–382 discussion 382–383
19. Elixhauser A, Steiner C, Fraser I. Volume thresholds and hospital characteristics in the United States. Health Aff (Millwood) 2003;22:167–177
20. 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
21. Pitt HA, Murray KP, Bowman HM, et al. Clinical pathway implementation improves outcomes for complex biliary surgery. Surgery 1999;126:751–756discussion 756–758
22. Pearson SD, Kleefield SF, Soukop JR, et al. Critical pathways intervention to reduce length of hospital stay. Am. J. Med. 2001;110:175–180
23. Sesperez J, Wilson S, Jalaludin B, et al. Trauma case management and clinical pathways: prospective evaluation of their effect on selected patient outcomes in five key trauma conditions. J. Trauma 2001;50:643–639
24. Kim MH, Rachwal W, McHale C, et al. Effect of amiodarone +/− diltiazem +/− beta blocker on frequency of atrial fibrillation, length of hospitalization, and hospital costs after coronary artery bypass grafting. Am. J. Cardiol. 2002;89:1126–1128
25. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. J. A. M. A. 1996;275:841–846
26. Khuri SF, Daley J, Henderson W, et al. 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
27. Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. Ann. Rev. Med. 2001;52:275–287
28. Hannan EL, Radzyner M, Rubin D, et al. The influence of hospital and surgeon volume on in-hospital mortality for colectomy, gastrectomy, and lung lobectomy in patients with cancer. Surgery 2002; 131:6–15
29. Cebul RD, Snow RJ, Pine R, et al. Indications, outcomes, and provider volumes for carotid endarterectomy. J. A. M. A. 1998;279:1282–1287
30. Hannan EL, O'Donnell JF, Kilburn H Jr., et al. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. J. A. M. A. 1989;262:503–510
31. Hannan EL, Popp AJ, Tranmer B, et al. Relationship between provider volume and mortality for carotid endarterectomies in New York State. Stroke 1998;29:2292–2297
32. Munoz E, Mulloy K, Goldstein J, et al. Costs, quality, and the volume of surgical oncology procedures. Arch. Surg. 1990;125:360–363
33. Christian CK, Gustafson ML, Betensky RA, et al. The leapfrog volume criteria may fall short in identifying high quality surgical centers. Ann. Surg. 2003;238:447–455
34. http://www.uhc.edu.