The Volume-Outcome Relationship: From Luft to Leapfrog

David M. Shahian, MD and Sharon-Lise T. Normand, PhD

Department of Thoracic and Cardiovascular Surgery, Lahey Clinic, Burlington, Massachusetts, Department of Health Care Policy, Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

Numerous reports have documented a volume-outcome relationship for complex medical and surgical care, although many such studies are compromised by the use of discharge abstract data, inadequate risk adjustment, and problematic statistical methodology. Because of the volume-outcome association, and because valid outcome measurements are unavailable for many procedures, volume-based referral strategies have been advocated as an alternative approach to health-care quality improvement. This is most appropriate for procedures with the greatest outcome variability between low-volume and high-volume providers, such as esophagectomy and pancreatectomy, and for particularly high-risk subgroups of patients. Whenever possible, risk-adjusted outcome data should supplement or supplant volume standards, and continuous quality improvement programs should seek to emulate the processes of high-volume, high-quality providers. The Leapfrog Group has established a minimum volume requirement of 500 procedures for coronary artery bypass grafting. In view of the questionable basis for this recommendation, we suggest that it be reevaluated.

(Ann Thorac Surg 2003;75:1048–58)

Numerous studies have demonstrated substantial variability in outcomes among providers of specialized medical and surgical care, including coronary artery bypass grafting (CABG) [1–5]. Reports from Canada, which has a greater regionalization of services and fewer low-volume providers, appear to show less such variation [6, 7]. In an attempt to explain these observations, many investigators have studied the structural characteristics of hospitals and individual practitioners, particularly their volume of procedures [2–4, 7–29].

The overall evidence supporting an association between volume and outcome is compelling. Dudley and colleagues [16, 30] reviewed 70 studies comparing the results for high-risk procedures at high-volume and low-volume hospitals. Sixty of these studies showed significantly better results at high-volume hospitals and 8 studies showed a nonsignificant trend toward better results. None of the studies demonstrated significantly worse results for high-volume providers. As part of an Institute of Medicine workshop report, Halm and associates [10] reviewed 88 of the most credible volume-outcome studies published since 1980. There was a statistically significant positive relationship between hospital or physician volume and better outcome in 77% of studies, including all 16 reports that were judged to have the best methodology. None of the studies showed a significant negative relationship. Based on available literature, Birkmeyer [20] estimated that high-volume providers have mortality rates that are 20% to 50% lower than low-volume providers for procedures such as CABG, carotid endarterectomy, and abdominal aortic aneurysm repair, and perhaps even greater differences for pancreatectomy and esophagectomy.

Volume-based referral strategies have been advocated by the Leapfrog Group and others, not only because of evidence supporting a volume-outcome relationship, but also because other options for improving quality have been regarded as impractical or ineffective [30]. Other structural measures of quality are not as predictive of outcome as volume [20]; process measures are not well understood; regulatory approaches are currently viewed with less enthusiasm [26]; and market-based approaches such as report cards have not been successful in influencing consumer behavior [31–34]. For infrequently performed procedures such as esophagectomy and pancreatectomy, it is also difficult to accrue sufficient patients for meaningful statistical analysis [24, 26, 35]. A continuous quality improvement approach can be difficult, costly, time-consuming, and has rarely been implemented [20, 23, 30]. Notably, both continuous quality improvement and report cards have been used extensively and effectively only in cardiac surgery [21, 34, 36–38].

In contrast to these other alternatives, Dudley and Johansen [30] assert that volume-based referral strategies are an immediate and feasible approach to quality improvement that will result, on average, in better outcomes. Volume thresholds may be used to encourage (market-based) or mandate (selective contracting) consumer use of designated hospitals, to justify restrictive licensure and certification of referral centers, to control the diffusion of new services, and to guide professional training requirements [17, 18, 20, 22, 23, 26, 30].

In this review, we present the evidence for a volume-outcome relationship in a variety of medical and surgical

Address reprint requests to Dr Shahian, Department of Thoracic and Cardiovascular Surgery, Lahey Clinic, 41 Mall Rd, Burlington, MA 01805; e-mail: david.m.shahian@lahey.org.

Published by Elsevier Science Inc

0003-4975/03/$30.00
PII S0003-4975(02)04308-4

conditions, examine some conceptual, methodological and statistical concerns regarding these studies, and consider the practical implications of regionalization. CABG studies are discussed in detail with particular attention to the 500-procedure volume threshold established by the Leapfrog Group.

Evidence Supporting the Volume-Outcome Relationship

Modern investigation of the volume-outcome hypothesis began with the work of Luft and colleagues [14] in 1979. These authors examined 1974 to 1975 discharge data on 12 surgical procedures from 1,498 hospitals. For certain operations (CABG, open heart surgery, vascular surgery, and TURP), hospitals that performed more than 200 procedures annually had patient-mix-adjusted death rates that were 25% to 41% lower than hospitals with fewer than 200 procedures (5.7% vs 3.4% for CABG). For a second group of procedures (colectomy, biliary tract surgery, total hip replacement, abdominal aortic aneurysm repair, vagotomy and pyloroplasty for ulcer disease, and cholecystectomy with common duct exploration), mortality stabilized at a much lower volume, ranging between 10 and 50 procedures per year. Finally, for vagotomy and cholecystectomy there was no observed volume-outcome relationship. In this prescient work, the authors identified many of the principles and questions that have been the basis of subsequent investigations over the next two decades, including (1) the applicability of the relationship to only more complex procedures; (2) procedure-specific functional relationships and thresholds; (3) the relative role of surgeon versus hospital volume; (4) the uncertain causality of the relationship; (5) the potential risks and benefits of regionalization; and (6) the need for comprehensive, risk-adjusted outcome data.

Subsequent studies have confirmed that the volume-outcome relationship persists despite adjustments for other structural characteristics [11, 12] and that there is considerable overlap in performance between high-volume and low-volume providers [39]. The volume-outcome association has been observed for both hospitals and individual providers, although some believe the evidence is stronger for the former [8, 25, 40]. The proportion of patients at an institution performed by low-volume providers is also important [13], and high-volume hospitals may actually enhance the performance of low-volume physicians [3, 41–43]. There remain many unanswered questions regarding the importance of the overall operative experience of surgeons, performance of closely related procedures, contemporaneous practice at more than one hospital, initial acquisition versus maintenance of skills, and transient versus permanent low-volume operators [13, 21, 23, 40, 44].

In the largest study currently available, Birkmeyer and associates [24] reviewed the results of 2.5 million Medicare patients who underwent 1 of 14 cancer or cardiovascular procedures between 1994 and 1999. Mortality and volume were inversely related for all these procedures, but the magnitude of the difference between very high-volume and very low-volume providers varied substantially (large differences for esophagectomy and pancreatectomy, small differences for CABG and carotid endarterectomy).

Certain specific conditions and procedures have been most extensively studied, including vascular surgery, cancer, and cardiac care. Many investigators [23, 41, 45–50], although not all [51–53], have documented lower postoperative death and stroke rates for patients undergoing carotid endarterectomy at high-volume hospitals. In a recent review by Birkmeyer and associates [24], of 479,289 Medicare patients, the difference between low-volume and high-volume hospitals was minimal (< 40 procedures, 1.7% mortality; > 164 procedures, 1.5% mortality). Studies of abdominal aortic aneurysm repair have generally demonstrated an inverse relationship of hospital volume and mortality for elective procedures [3, 45–47, 54–57] and surgeon volume and mortality for ruptured aneurysms [55, 58].

Hillner [59] reviewed volume-outcome studies in cancer treatment and concluded that the relationship existed primarily for more complex and less frequently performed procedures. Hannan and colleagues [60] used hierarchical models to study New York patients undergoing cancer resection of the colon, stomach, or lung between 1994 and 1997, and an inverse relationship between mortality and volume for surgeons and hospitals was noted. Begg and associates [61] reviewed the outcomes of 5,013 Medicare patients who underwent resections of the pancreas, esophagus, liver, lung, or pelvis between 1984 and 1993. There was a significant inverse relationship between volume and mortality for all procedures except pneumonectomy. For esophagectomy, 30-day mortality for the low-volume providers (1 to 5 procedures per year) and high-volume providers (> 11 procedures per year) was 17.3% and 3.4%, respectively, and for pancreatectomy the corresponding values were 12.9% and 5.8%, respectively. These two operations provide the most dramatic examples of the volume-outcome association, probably because of their relative infrequency, technical complexity, and challenging postoperative care [62–69]. In the recent Medicare study by Birkmeyer and colleagues [24], adjusted surgical mortality for pancreatectomy varied from 16.3% (< 1 procedure per year) to 3.8% (> 16 procedures per year), and for esophagectomy the mortality ranged from 20.3% (< 2 procedures per year) to 8.4% (> 19 procedures per year).

Cancer studies have also demonstrated the impact of volume on endpoints other than surgical mortality, including the use of more advanced surgical techniques, perioperative morbidity, local recurrence, and long-term survival [67, 70–74]. This reflects the importance not just of surgical skill but also process issues such as treatment selection, adjuvant therapy, and long-term follow-up.

The volume-outcome association also has been investigated in various types of cardiac care including heart transplant [75], pediatric cardiac surgery [57, 76, 77], myocardial infarction [78, 79], percutaneous transluminal coronary angioplasty [42, 43, 80–83], and CABG [2–4, 7, 14, 21, 28, 34, 38–40, 57, 84–86].

Ann Thorac Surg
2003;75:1048–58

Coronary Artery Bypass Grafting

CABG is a mature procedure that has been performed for three decades, and cardiac surgeons have led all other medical specialties in the development of risk-adjusted databases, publication of outcomes, and continuous quality improvement activities. CABG is also unique among high-complexity procedures in its frequency. Data presented by Birkmeyer and associates [17] suggest that CABG is 10 times as common as abdominal aortic aneurysm resection in the United States, 150 times as common as esophagectomy, and 2½ times as common as carotid endarterectomy. Ironically, despite being the most frequently performed and thoroughly investigated of all complex surgical operations, CABG has also been targeted with the most aggressive and restrictive volume standards.

Beginning with the original work of Luft and associates [14] in 1979, showing a volume threshold of about 200 procedures, CABG was a common focus of early volume-outcome studies [4, 28, 39]. Showstack and associates [84] studied 18,986 patients who underwent CABG at one of 77 California hospitals in 1983. Overall, hospitals that performed 20 to 100 procedures per year or 201 to 350 procedures per year had significantly worse outcomes than the highest volume hospitals (> 350 procedures per year). For scheduled operations, only hospitals with 201 to 350 procedures per year had statistically significantly higher mortality (2.9% vs 2.2%); however, for nonscheduled patients, hospitals with 20 to 100 or 201 to 350 procedures annually had significantly higher mortality. Hospitals that performed 101 to 200 CABG procedures per year, which accounted for 28% of total hospitals and 16% of all CABG patients, did not differ significantly from the highest volume group in either the scheduled or nonscheduled category, and there was only a marginal difference when all patients were considered ($p < 0.10$). The greatest lifesaving potential in this study appeared to be for nonscheduled patients, similar to the findings of Nallamothu and associates [87] that a significant mortality difference between low-volume (< 200 procedures annually) and high-volume hospitals was observed only in the minority of patients who were high-risk.

Grumbach and associates [7] used 1987 to 1989 discharge abstract data to compare the results of CABG operations in New York, Canada, and California. In New York and Canada, 60% of CABG procedures were performed in hospitals with an annual CABG volume exceeding 500 procedures, whereas the corresponding value for California was only 26%. Four percent of New York and 2% of Canadian CABG procedures were performed in hospitals with an annual CABG volume of fewer than 200 procedures per year, whereas over one-third of the patients in California were performed in such hospitals. The highest mortality rate in this study was in California hospitals performing fewer than 100 CABG procedures per year, which comprised 31% of the California hospitals. Interestingly, just as in the study of Showstack and associates [84], the mortality for hospitals performing 100 to 199 annual procedures in New York was not significantly different from hospitals that performed more than 500 procedures per year ($p = 0.16$). Similarly, mortality for the 200 to 499 procedures per year Canadian group was no different than that of Canadian hospitals performing more than 500 procedures per year ($p = 0.51$).

In the most recent and largest volume-outcome study of CABG based on administrative data, Birkmeyer and associates [24] have reported the results of 901,667 Medicare patients who underwent CABG at 1,068 hospitals between 1994 and 1999. The range of adjusted mortality was much smaller than in earlier Medicare studies, varying from 5.6% (< 230 estimated total CABG procedures per hospital) to 4.5% (> 849 procedures). Over the broad middle range of volume from 230 procedures to 849 procedures, mortality varied only from 5.1% to 4.7%. This may be an important demonstration of the decreasing impact of volume over time as procedures mature.

In 1991, Hannan and colleagues [2] reported the first volume-outcome study for CABG to use properly risk-adjusted clinical data. This study has been cited as one of the best in the literature by two principal advisors [16, 17] to the Leapfrog group on evidence-based referrals, and it has undoubtedly influenced the recommendation of that organization regarding volume standards for CABG. The data source for this study included 12,448 patients who underwent CABG in New York during 1989, which was the first year of the New York Cardiac Surgery Reporting System (CSRS). Higher volume hospitals had higher expected mortality, which contradicts the findings of some other studies [11, 12, 14, 87, 88] that such hospitals may attract better risk patients. Using a logistic regression model, the log of surgeon volume was significantly related to mortality ($p < 0.001$), but hospital volume was only "marginally related" to mortality ($p = 0.04$). Risk-adjusted hospital mortality ranged from 7.25% in hospitals performing between 1 to 199 procedures per year to 2.85% for hospitals performing more than 890 cardiac procedures per year. Risk-adjusted mortality for surgeons performing fewer than 55 procedures per year was 8.14% compared with 2.43% for surgeons performing 260 or more procedures per year. High-volume surgeons in high-volume hospitals achieved the lowest mortality (2.18%).

Several important aspects of this study must be emphasized. First, in contrast to virtually every other publication relating CABG volume and mortality, these authors used total number of cardiac operations as the volume indicator, which is in our opinion a much better measure of overall cardiac surgical experience. However, volumes cited in this article must be interpreted with caution because some refer to total cardiac experience and some to CABG only. Second, the data in this study came from 1989, the first year of data collection in the New York Cardiac Surgery Reporting System. This was the least reliable year of the New York database, primarily because of under-coding and over-coding of comorbidities [31, 89].

Finally, the nature of the functional relationship between volume and mortality in this and other studies

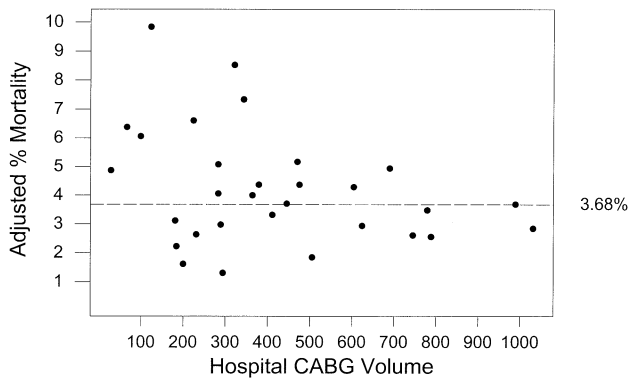

Fig 1. Hospital coronary artery bypass grafting (CABG) *volume and adjusted mortality for New York state in 1989 [2].*

must be closely scrutinized. We extracted hospital mortality and volume data from this article, and we present them in graphical form in Figures 1 to 3. Figure 1 is a scatterplot of the adjusted mortality rate versus CABG volume for each hospital. At higher volume levels, there is less variability, although a few programs exceed the state average risk-adjusted mortality. Over the broad middle range of volume, there is increasing dispersion of results on both sides of the state average as sample sizes decrease and confidence intervals broaden, consistent with the behavior of a binomial proportion. Importantly, numerous hospitals performing 150 to 500 procedures per year had risk-adjusted mortality rates well below the state average. The outcomes in the very lowest volume group of hospitals are clearly anomalous and skewed toward high mortality; their performance suggests that they may have a different functional relationship between volume and mortality than the remainder of hospitals. Figure 2 is a plot of the log odds of observed mortality versus CABG volume. The horizontal line is the log odds of observed mortality for the entire state, and the irregular line is a LOWESS smoother (robust locally

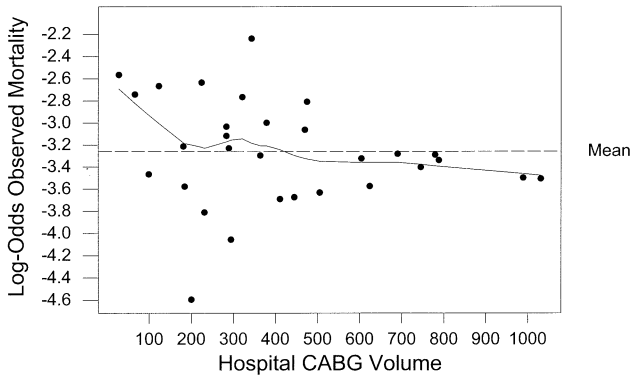

Fig 2. Hospital coronary artery bypass grafting (CABG) *volume and log-odds of observed hospital mortality for New York state in 1989, as derived from the data of Hannan and associates [2]. The* curvilinear line *is a LOWESS scatterplot smoother showing the shape of the relationship between the two variables. The* horizontal line *is the log-odds of observed mortality for the entire state.*

weighted scatterplot smoother; f = 0.5, 2 iterations; MINITAB Statistical Software, Release 13 for Windows; State College, PA), a nonparametric estimate of the shape of the relationship between these variables [90]. Figure 3 is the same plot for the log-odds adjusted mortality.

Dudley and associates [16] derived a 500-case threshold from this study by Hannan and colleagues [2] and used this threshold to estimate lives saved by regionalization. This recommendation is also likely to have been the basis for the Leapfrog group 500-case standard for CABG [17]. However, nowhere in the original article by Hannan and colleagues [2] is such a threshold mentioned, and visual inspection of our graphs derived from these data (Figs 1 – 3) suggests that the most significant mortality change point occurs at a much lower volume of less than 300 procedures. The Society of Thoracic Surgeons has performed an independent analysis of these data and identified a threshold of approximately 125 procedures per year [40].

Contrary Evidence

The most compelling evidence against a general volume-outcome relationship comes from the Veterans Affairs Administration [51, 91]. Using a comprehensive risk-adjusted database and a variety of statistical techniques, Khuri and associates [51] investigated outcomes for eight common, noncardiac procedures among 68,631 patients operated upon in Veterans Affairs (VA) hospitals. No relationship of volume to 30-day mortality was noted, and there was no volume threshold. Although these results appear to conflict with those of most other studies, there are significant caveats regarding generalizations from VA data [30, 51, 85, 92]. The patients are unusually homogenous and primarily older males. VA hospitals are generally low to moderate volume, and thus the range of patient volumes available for comparative analysis is somewhat restricted. Several procedures (esophagectomy and pancreatectomy) having the strongest volume-outcome association in prior studies were not included in the report of Khuri and associates [51]. In the VA, many processes and structural requirements have already been standardized and results have been monitored for years. Finally, there is a close affiliation of many VA hospitals with high-volume academic centers, and they may consequently benefit from common staff, processes, and oversight procedures.

With regard to the volume-outcome relationship for CABG, Clark [86] reviewed data from 1991 to 1993 on 120,377 patients from 180 practice groups (not hospitals) that had submitted their results to the Society of Thoracic Surgeons database. Procedural volumes ranged from fewer than 100 per year to more than 900 per year. Linear and logistic regression showed a "weak to very weak inverse correlation" between volume and mortality (r^2 = 0.0492; odds ratio = 1.000), but the only "clinically relevant" correlation of volume to mortality was for groups performing fewer than 100 procedures per year. Notably, 56.1% of patients were operated on by groups performing 500 or fewer procedures per year, which is the current volume standard of the Leapfrog group. Arguably, the

Ann Thorac Surg
2003;75:1048–58

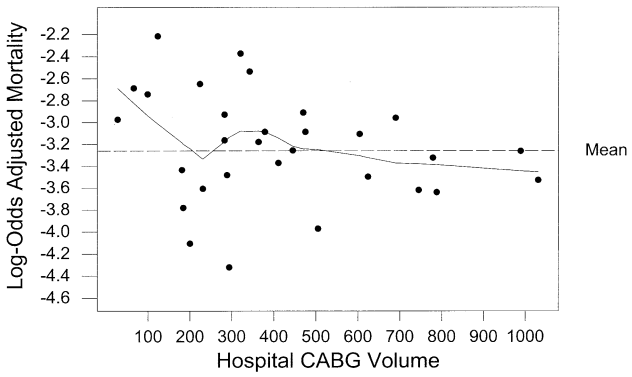

*Fig 3. Hospital coronary artery bypass grafting (*CABG*) volume and log-odds of adjusted mortality for New York state in 1989 [2]. The* curvilinear line *is a LOWESS scatterplot smoother showing the shape of the relationship between the two variables. The* horizontal line *is the log-odds of observed mortality for the entire state.*

Society of Thoracic Surgeons database is voluntary and may be skewed toward groups with better results. Also this study focuses on groups, not hospitals or surgeons, and thus it is not strictly comparable with most other volume-outcome studies.

Shroyer and colleagues [85] investigated the results of 23,986 CABG procedures performed between 1987 and 1992 at 44 VA hospitals. When hospital CABG volume was used as a covariate in a multivariate risk model, it had no relationship to mortality. In hospital models, threshold analysis demonstrated a change point at 100 CABG procedures per year, although this was not confirmed in a Poisson additive model. Although this study is large and well documented, it may not be generalizable because of the same caveats discussed previously regarding the unique features of VA hospitals.

Sollano and associates [57] studied discharge abstract data of 97,137 New York patients undergoing CABG between 1990 and 1995 (the period after implementation of a statewide report card). There was no relationship of volume to mortality for CABG, although there was a relationship of volume to mortality for pediatric cardiac surgery, and the authors attributed this difference to the impact of public outcome reporting of CABG.

Controversial Aspects of the Volume-Outcome Hypothesis

Despite the preponderance of studies that support a volume-outcome relationship, and the increasing use of this hypothesis to justify major health policy decisions, many unanswered questions remain.

Conceptual

A certain level of training and repetition is important for any complex task, and many studies document a continuous decline in surgical mortality with increasing volume. However, common experience in nonmedical areas would suggest that volume is not an absolute prerequisite for high quality, nor is it necessarily even the most frequently used criterion. For specialized medical care, some threshold level of volume is required to develop the technical proficiency and processes necessary for optimal patient care. However, notwithstanding this caveat, high-quality results are often achieved by low to moderate-volume providers, and high volume does not guarantee high quality.

Methodology of Volume-Outcome Studies

DATA SOURCE. Although a few volume-outcome studies are based on clinical databases [2], the majority are limited by their use of administrative data from discharge abstracts [23]. Many studies focus only on the Medicare population because of the availability of that data [24]. Birkmeyer [18], a proponent of volume-based referral, also acknowledges that many studies are skewed by data from both high-volume referral centers and from low-volume California hospitals, and this may limit their general applicability. Furthermore, many reports are outdated and do not take into account advancements in medical and surgical care [24]. The performance gap between low-volume and high-volume providers has been observed to narrow over time [23, 83, 93] because of both general advancements in medicine and the maturation of specific procedures.

RISK ADJUSTMENT. The degree of risk adjustment has varied widely among volume-outcome studies [23]. Sowden and associates [93] reviewed seven publications based on United States CABG results from 1972 to 1992. As risk adjustment improved from worst to best, the odds ratio for mortality at high-volume hospitals narrowed from 0.54 to 0.84. The mitigating effect of risk-adjusted outcomes on the magnitude of the volume-outcome relationship has been also noted in numerous other studies [47, 85, 94], but it is unlikely that this omission would have uniformly favored high-volume hospitals [30], or that it would change the general nature of the observed relationship [23].

CROSS-SECTIONAL VERSUS LONGITUDINAL ANALYSIS. Another methodological concern is the cross-sectional nature of most studies. Comparisons are made during one time period among many hospitals, often differing in location, structural characteristics, patient numbers, and severity. As Hannan [22] has observed, a better way to test the relationship between volume and outcome would be a longitudinal study, assessing what happens at a single institution over time as its volume changes. This has the advantage of holding constant many other characteristics of the hospital and its patient population. Few investigations [55, 95] have studied such longitudinal relationships because provider volumes rarely undergo dramatic changes over time. However, in a recent longitudinal study using hierarchical models, Bronskill and colleagues [96] found no relationship between surgeon volume for CABG and in-hospital mortality using 1991 to 1995 data from the New York State Cardiac Surgery Reporting System.

STATISTICAL ANALYSIS. Statistical methodologies have varied widely among volume-outcome studies and include standard multivariable regression and logistic regression, Poisson additive models [85], and hierarchical models [51, 60, 96]. Most studies have modeled the relationship between volume and mortality as a single, continuous, nonlinear function (logarithmic, exponential decay, quadratic, and hyperbolic). The functional relationship appears to vary from procedure to procedure, as do the domain and range of the variables and the inflection points of their volume-mortality curves. The performance of a few extremely low-volume providers is largely responsible for the significant results in many studies. If these were considered as a separate aberrant group rather than being forced into a global functional relationship with the remaining hospitals, such studies would likely demonstrate a less significant volume-outcome association over the intermediate range of volumes.

Volume thresholds, such as the 500-procedure standard established for CABG by the Leapfrog Group, imply two distinct aspects of the relationship between hospital volume and mortality. First, they suggest that the relationship is different for hospitals performing fewer than 500 procedures per year than those performing more than this number. More importantly, such thresholds also imply that hospitals within each of these volume groups have a similar volume-mortality relationship. This may be true of hospitals performing many procedures where we expect an asymptote, but for smaller volume there may be a decreasing relationship as volume increases. A more realistic approach to modeling the volume-outcome relationship should incorporate spline functions (piecewise polynomial functions, often higher order) or other flexible parameterizations that can incorporate different functional relationships for each volume interval.

Hospital or physician volume, the predictor variable of interest in volume-outcome studies, has been analyzed as either a continuous or categorical variable. In the latter, a change point is selected and providers are categorized into low-volume or high-volume groups for analysis. These change points have been variously determined from external literature standards, visual inspection of volume-outcome curves, mean or median volume, sequential analysis of variance, and Poisson additive models. Altman and colleagues [97] have discussed the disadvantages of categorizing continuous data in this fashion, including not only the loss of important information but also the potential for introducing Type I statistical error. As noted by Halm and associates [10], it has not even been determined whether it is more appropriate to group together all providers performing in excess of some low-volume threshold, or if only the highest volume providers have the best results. Finally, the definition of low volume varies dramatically in the literature depending upon the procedure being studied. This designation has been applied to providers performing fewer than 5 to 10 procedures per year for certain operations like esophagectomy and pancreatectomy, whereas for CABG it has often included hospitals performing hundreds of procedures per year. The Leapfrog Group recommendations imply that anything less than 500 CABG procedures per year is low volume.

These considerations are of more than theoretical statistical interest. Although the determination of volume thresholds appears somewhat arbitrary in many published studies, these have been extracted and used by insurers and regulators to impose minimum volume requirements. The lack of a formal statistical approach to the identification and estimation of the volume change point is one of the most troubling aspects of the volume-outcome debate. A principled approach to this inferential question, characterized in the statistical literature as an inverse prediction or a change point problem, should involve not only the estimation of the change point but also its 95% confidence interval. An example of such an approach is the determination of optimal body mass index (BMI) from data on the relationship between BMI and mortality [98]. Health policy planners must decide their primary goal in studying the volume-outcome relationship. Is it to exclude those providers expected to be the lowest performing in a region or selectively refer only to those providers predicted to have the highest performance (both of which are threshold or change point problems)? Or is it to predict the level of volume that will achieve some predetermined outcome standard (an inverse prediction problem)? Once these objectives are determined, the appropriate statistical techniques, which are frequently complex and nontrivial, can be applied.

In most volume-outcome studies the endpoint (typically mortality) has a low incidence. By the very nature of the study design, the sample sizes differ substantially among providers and are sometimes quite small; this has a considerable impact upon the reliability of interprovider comparisons using classic inferential statistics [26, 35]. Typically, a sample of size n is taken from a provider and is used to make inferences about its true underlying population proportion (ie, mortality), which is the real value of interest. The larger the sample-n, the more certain we are that the sample proportion is a good estimate of the population proportion. Conversely, with small sample sizes we have less confidence about the true performance of the provider, and much of the observed increase in variability may be random. Because of small sample sizes, the observed results of lower-volume providers are greatly affected by even one or two deaths [38]. In addition to the greater statistical uncertainty of estimates when sample sizes are small, Stein's theorem [99] also suggests that a better estimate of the true performance of such low-volume providers may be obtained by shrinking their observed mortality rates toward the mean of like providers. This reduces the possibility they will be falsely labeled as outliers, one of many advantageous features of hierarchical models [32, 100, 101]. For all these reasons, investigators regard small sample size as one of most troubling methodological weaknesses of volume-outcome research [18, 22, 26, 35, 88], just as it is also a limiting factor in producing credible risk-adjusted outcome reports [32, 102].

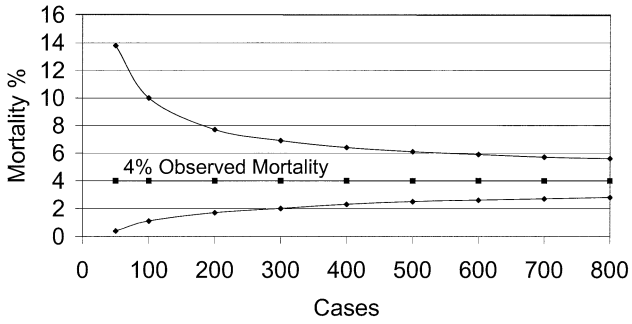

*Fig 4. Exact 95% confidence intervals (*diamonds*) for 4% observed mortality (*squares*) at varying procedure volumes (sample sizes). The confidence intervals widen substantially as sample sizes decrease, making valid interprovider comparisons problematic using standard inferential statistical techniques.*

This statistical concept is illustrated in Figure 4, which depicts the hypothetical results of nine hospitals with annual caseloads varying from 50 to 800 procedures, all of which have a 4% observed mortality rate. For hospitals having lower yearly volumes, the exact 95% binomial confidence intervals (MINITAB Statistical Software, Release 13 for Windows, State College, PA) of these estimates broaden dramatically, thus indicating that they are a less certain measure of the true performance of these hospitals.

Other features of volume-outcome studies also present problems for classic statistical analysis. Provider characteristics are highly correlated, and this limits the confidence with which researchers can claim that procedural volume is related to mortality. For example, high-volume hospitals are more likely to be teaching hospitals in urban areas. Isolating the effect of volume, even with sophisticated statistical models, is challenging at best. There also may be clustering of high-risk and low-risk patients within a provider, which violates the assumption of independence. There are multiple levels of aggregation (patient, surgeon, hospital, low-volume provider and high-volume provider groups), and unless hierarchical models are used, this results in underestimation of the random component of variation at each level and overestimation of systematic variation [32, 96, 100, 101]. In studies of small-area variation, use of hierarchical models resulted in less apparent systematic variation than standard analyses [103].

Causality

Because the volume-outcome relationship has increasingly become the basis for major health policy recommendations, it is troubling that its mechanism remains problematic. Historically, the two major hypotheses are *practice makes perfect* and *selective referral* [55, 95, 104]. In the former, it is presumed that increasing volume brings additional expertise, whereas the latter formulation suggests that consumers are drawn to hospitals and physicians with perceived or historically high performance.

It is likely that processes of care also may be largely responsible for the volume-outcome relationship. Research efforts should focus on identifying and disseminating these processes, which may cause less disruption to the health care system than shifting patients among providers [18, 23]. Possible process factors include appropriate selection of patients and treatment type, integrated perioperative care, adjuvant therapy, and long-term follow-up [18, 30, 62, 67, 68, 70, 105].

These various hypotheses have more than theoretical importance. If practice makes perfect is the mechanism, then it should be possible to channel large numbers of patients to almost any institution with the expectation that increased volume will produce better outcomes. However, if selective referral is the mechanism, then contracting or directing patients on any criteria other than actual current performance (eg, low price) may result in unchanged or even poorer outcomes [26, 39, 55, 104]. In this context, it is important to note that the only two reports that specifically studied this question for CABG identified selective referral as the mechanism [95, 104].

Practical Implications of Regionalization

The volume-outcome relationship has been used to justify regionalization of specialty services. This is often based upon estimates of lives saved, a statistical extrapolation that typically requires substantial assumptions. Tempering the enthusiasm for regionalization are a number of potential negative consequences [16–20, 23, 26–28, 30, 40, 106] for the *patient* (delay or avoidance in seeking care, loss of continuity with primary physician, preference for local care regardless of outcomes [107], small possibility of inferior care at referral center [23]), for *low-volume hospitals* (unfair stigmatization, closure of related services, financial viability, incentive to inappropriate utilization), for *high-volume hospitals* (inability to sufficiently increase capacity and maintain quality), and for the *regional health care market* (economic distortion, unmerited halo effect and excessive contractual power for high-volume providers, diminished incentive to develop risk-adjusted outcome reporting and continuous quality improvement programs).

Some studies [27, 28] demonstrate that full regionalization (ie, sending all patients to the highest volume hospitals) would disenfranchise a high percentage of institutions currently performing complex procedures (eg, 80% of urban hospitals currently performing elective abdominal aortic aneurysm repair [44]), and this would simultaneously create an increased level of demand at high-volume hospitals that may be impossible to accommodate.

The practical consequences of redirecting large numbers of patients to high-volume providers has led some to propose selective or partial regionalization. This would be based on the mortality spread between low-volume and high-volume providers, the number of patients undergoing a particular procedure, patient acuity, and practical considerations [16, 17, 19, 27, 28, 30, 87]. As suggested by Epstein [44], the least disruption to the health care system might result from targeting procedures with the greatest mortality difference between low-volume and

high-volume providers, even though some more frequently performed procedures with a smaller difference in outcome may potentially impact a larger number of patients. Based upon the CABG studies of Showstack [84] and Nallamothu [87], targeted regionalization for particularly high-risk subsets of patients also should be considered.

Conclusion

Volume-based referral strategies are best used for procedures like esophagectomy and pancreatectomy with large differences in outcomes between low-volume and high-volume providers, and also for particularly high-risk patient subsets. For other complex procedures with a less dramatic range of mortality, volume standards may be used as a red flag to focus attention on the lowest volume providers, not because their results are uniformly poor but because their small numbers make it difficult to accurately assess their performance. Whenever possible, risk-adjusted outcome data should be used to supplement or supplant such volume standards [22]. Continuous quality improvement programs should study and then seek to emulate as many of the process and structural characteristics of high-volume, high-quality providers that are feasible.

Although the Leapfrog Group volume thresholds for other invasive procedures are generally compatible with published outcome data, their 500-procedure volume standard for CABG is inconsistent with the results of empirical studies and the recommendations of major professional societies. Extensive reviews by the Society of Thoracic Surgeons [40, 86] and the Veterans Affairs Administration [85] demonstrate no significant volume effect except at very low levels (100 procedures per year), and the 1999 joint guidelines of the American Heart Association and American College of Cardiology only recommend close monitoring of programs performing fewer than 100 procedures per year [108]. The American College of Surgeons suggests that 100 to 125 procedures per year are sufficient to maintain quality, but that at least 200 procedures are necessary to function efficiently [109].

At a time when overall CABG rates are declining because of percutaneous transluminal coronary angioplasty, establishment of a 500-procedure standard for CABG may encourage inappropriate utilization, disenfranchise a substantial number of high-quality programs, threaten teaching programs, or adversely impact angioplasty and other urgent cardiac care. These volume thresholds should be reevaluated using more appropriate statistical methods for inferring change points from the empirical data, and the Leapfrog Group should also adhere to its plan to waive volume standards whenever high-quality outcome data are available.

We can provide no better conclusion than this recent caveat from Professor Luft [88], the pioneer in volume-outcome research:

"Given that volume per se does little more than assure reasonable confidence intervals around statistical estimates, volume-based policies for excluding or closing facilities are inferior to policies based on the routine reporting of risk-adjusted outcomes and policies to encourage the referral of patients toward facilities with better than expected outcomes and away from those with worse than expected outcomes."

References

1. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. The Northern New England Cardiovascular Disease Study Group. JAMA 1991; 266:803–9.
2. Hannan EL, Kilburn H Jr, Bernard H, O'Donnell JF, Lukacik G, Shields EP. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. Med Care 1991;29:1094–107.
3. Hannan EL, O'Donnell JF, Kilburn H Jr, Bernard HR, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. JAMA 1989;262:503–10.
4. Luft HS, Romano PS. Chance, continuity, and change in hospital mortality rates. Coronary artery bypass graft patients in California hospitals, 1983 to 1989. JAMA 1993;270: 331–7.
5. Williams SV, Nash DB, Goldfarb N. Differences in mortality from coronary artery bypass graft surgery at five teaching hospitals. JAMA 1991;266:810–5.
6. Tu JV, Naylor CD. Coronary artery bypass mortality rates in Ontario. A Canadian approach to quality assurance in cardiac surgery. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. Circulation 1996; 94:2429–33.
7. Grumbach K, Anderson GM, Luft HS, Roos LL, Brook R. Regionalization of cardiac surgery in the United States and Canada. Geographic access, choice, and outcomes. JAMA 1995;274:1282–8.
8. Kelly JV, Hellinger FJ. Physician and hospital factors associated with mortality of surgical patients. Med Care 1986; 24:785–800.
9. Halm EA, Chassin MR. Why do hospital death rates vary? N Engl J Med 2001;345:692–4.
10. Halm EA, Lee C, Chassin MR. Appendix C. How is volume related to quality in health care? A systematic review of the research literature. In: Hewitt M, editor. Interpreting the volume-outcome relationship in the context of health care quality: workshop summary. Washington, DC: National Academy Press, 2000.
11. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: The relationship between hospital volume and outcomes for selected diagnostic categories. Med Care 1984;22:98–114.
12. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: The relation between volume and outcomes and other hospital characteristics. Med Care 1984;22:115–25.
13. Hughes RG, Hunt SS, Luft HS. Effects of surgeon volume and hospital volume on quality of care in hospitals. Med Care 1987;25:489–503.
14. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 1979;301:1364–9.
15. Luft HS, Garnick DW, Mark DH. Hospital volume, physician volume, and patient outcomes: assessing the evidence. Ann Arbor: Health Administration Press, 1990.
16. Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000;283:1159–66.
17. Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume stan-

dards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. Surgery 2001;130:415–22.
18. Birkmeyer JD. High-risk surgery–follow the crowd. JAMA 2000;283:1191–3.
19. Birkmeyer JD, Lucas FL, Wennberg DE. Potential benefits of regionalizing major surgery in Medicare patients. Eff Clin Pract 1999;2:277–83.
20. Birkmeyer JD. Should we regionalize major surgery? Potential benefits and policy considerations. J Am Coll Surg 2000;190:341–9.
21. Hannan EL, Siu AL, Kumar D, Kilburn H Jr, Chassin MR. The decline in coronary artery bypass graft surgery mortality in New York State. The role of surgeon volume. JAMA 1995;273:209–13.
22. Hannan EL. The relation between volume and outcome in health care. N Engl J Med 1999;340:1677–9.
23. Hewitt M. Interpreting the volume-outcome relationship in the context of health care quality: workshop summary. Washington, DC, National Academy Press: 1–18;2000.
24. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. N Engl J Med 2002;346:1128–37.
25. Houghton A. Variation in outcome of surgical procedures. Br J Surg 1994;81:653–60.
26. Phillips KA, Luft HS. The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment. Int J Qual Health Care 1997;9:341–8.
27. Sloan FA, Perrin JM, Valvona J. In-hospital mortality of surgical patients: is there an empiric basis for standard setting? Surgery 1986;99:446–54.
28. Maerki SC, Luft HS, Hunt SS. Selecting categories of patients for regionalization. Implications of the relationship between volume and outcome. Med Care 1986;24:148–58.
29. Riley G, Lubitz J. Outcomes of surgery among the Medicare aged: surgical volume and mortality. Health Care Financ Rev 1985;7:37–47.
30. Dudley RA, Johansen KL. Invited commentary: physician responses to purchaser quality initiatives for surgical procedures. Surgery 2001;130:425–8.
31. Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. N Engl J Med 1996;334:394–8.
32. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. Ann Thorac Surg 2001;72:2155–68.
33. Shahian DM, Yip W, Westcott G, Jacobson J. Selection of a cardiac surgery provider in the managed care era. J Thorac Cardiovasc Surg 2000;120:978–87.
34. Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. Ann Thorac Surg 1994;58:1852–7.
35. Luft HS, Hunt SS. Evaluating individual hospital quality through outcome statistics. JAMA 1986;255:2780–4.
36. 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. JAMA 1996;275:841–6.
37. Grover FL, Johnson RR, Shroyer AL, Marshall G, Hammermeister KE. The Veterans Affairs Continuous Improvement in Cardiac Surgery Study. Ann Thorac Surg 1994;58:1845–51.
38. Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. JAMA 1994;271:761–6.
39. Freeland MS, Hunt SS, Luft HS. Selective contracting for hospital care based on volume, quality, and price: prospects, problems, and unanswered questions. J Health Polit Policy Law 1987;12:409–26.
40. Crawford FA Jr, Anderson RP, Clark RE, et al. Volume requirements for cardiac surgery credentialing: a critical examination. The Ad Hoc Committee on Cardiac Surgery Credentialing of The Society of Thoracic Surgeons. Ann Thorac Surg 1996;61:12–6.
41. Cebul RD, Snow RJ, Pine R, Hertzer NR, Norris DG. Indications, outcomes, and provider volumes for carotid endarterectomy. JAMA 1998;279:1282–7.
42. Hannan EL, Racz M, Ryan TJ, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;277:892–8.
43. Hannan EL. Percutaneous coronary interventions: heed the American College of Cardiology volume recommendations, but strive to improve quality. J Am Coll Cardiol 1999;34:1481–3.
44. Epstein AM. Volume and outcome–it is time to move ahead. N Engl J Med 2002;346:1161–4.
45. Manheim LM, Sohn MW, Feinglass J, Ujiki M, Parker MA, Pearce WH. Hospital vascular surgery volume and procedure mortality rates in California, 1982–1994. J Vasc Surg 1998;28:45–56.
46. Pearce WH, Parker MA, Feinglass J, Ujiki M, Manheim LM. The importance of surgeon volume and training in outcomes for vascular surgical procedures. J Vasc Surg 1999;29:768–76.
47. Shackley P, Slack R, Booth A, Michaels J. Is there a positive volume-outcome relationship in peripheral vascular surgery? Results of a systematic review. Eur J Vasc Endovasc Surg 2000;20:326–35.
48. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. JAMA 1998;279:1278–81.
49. Hannan EL, Popp AJ, Tranmer B, Fuestel P, Waldman J, Shah D. Relationship between provider volume and mortality for carotid endarterectomies in New York state. Stroke 1998;29:2292–7.
50. Huber TS, Seeger JM. Dartmouth Atlas of Vascular Health Care review: impact of hospital volume, surgeon volume, and training on outcome. J Vasc Surg 2001;34:751–6.
51. Khuri SF, Daley J, Henderson W, et al. Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. Ann Surg 1999;230:414–29.
52. Kirshner DL, O'Brien MS, Ricotta JJ. Risk factors in a community experience with carotid endarterectomy. J Vasc Surg 1989;10:178–86.
53. Kempczinski RF, Brott TG, Labutta RJ. The influence of surgical specialty and caseload on the results of carotid endarterectomy. J Vasc Surg 1986;3:911–6.
54. Katz DJ, Stanley JC, Zelenock GB. Operative mortality rates for intact and ruptured abdominal aortic aneurysms in Michigan: an eleven-year statewide experience. J Vasc Surg 1994;19:804–15.
55. Hannan EL, Kilburn H Jr, O'Donnell JF, et al. A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. Health Serv Res 1992;27:517–42.
56. Wen SW, Simunovic M, Williams JI, Johnston KW, Naylor CD. Hospital volume, calendar age, and short term outcomes in patients undergoing repair of abdominal aortic aneurysms: the Ontario experience, 1988–92. J Epidemiol Community Health 1996;50:207–13.
57. Sollano JA, Gelijns AC, Moskowitz AJ, et al. Volume-outcome relationships in cardiovascular operations. New York State, 1990–1995. J Thorac Cardiovasc Surg 1999;117:419–28.
58. Dardik A, Burleyson GP, Bowman H, et al. Surgical repair of ruptured abdominal aortic aneurysms in the state of Maryland: factors influencing outcome among 527 recent cases. J Vasc Surg 1998;28:413–20.
59. Hillner BE. Is cancer care best at high-volume providers? Curr Oncol Rep 2001;3:404–9.
60. Hannan EL, Radzyner M, Rubin D, Dougherty J, Brennan MF. 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.
61. Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. JAMA 1998;280:1747–51.
62. Gordon TA, Burleyson GP, Tielsch JM, Cameron JL. The effects of regionalization on cost and outcome for one general high-risk surgical procedure. Ann Surg 1995;221: 43–9.
63. Lieberman MD, Kilburn H, Lindsey M, Brennan MF. Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 1995;222:638–45.
64. Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. West J Med 1996;165:294–300.
65. Sosa JA, Bowman HM, Gordon TA, et al. Importance of hospital volume in the overall management of pancreatic cancer. Ann Surg 1998;228:429–38.
66. Birkmeyer JD, Finlayson SR, Tosteson AN, Sharp SM, Warshaw AL, Fisher ES. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. Surgery 1999;125:250–6.
67. Birkmeyer JD, Warshaw AL, Finlayson SR, Grove MR, Tosteson AN. Relationship between hospital volume and late survival after pancreaticoduodenectomy. Surgery 1999; 126:178–83.
68. Dimick JB, Cattaneo SM, Lipsett PA, Pronovost PJ, Heitmiller RF. Hospital volume is related to clinical and economic outcomes of esophageal resection in Maryland. Ann Thorac Surg 2001;72:334–9.
69. Kuo EY, Chang Y, Wright CD. Impact of hospital volume on clinical and economic outcomes for esophagectomy. Ann Thorac Surg 2001;72:1118–24.
70. Roohan PJ, Bickell NA, Baptiste MS, Therriault GD, Ferrara EP, Siu AL. Hospital volume differences and five-year survival from breast cancer. Am J Public Health 1998;88: 454–7.
71. Hodgson DC, Fuchs CS, Ayanian JZ. Impact of patient and provider characteristics on the treatment and outcomes of colorectal cancer. J Natl Cancer Inst 2001;93:501–15.
72. Marusch F, Koch A, Schmidt U, Pross M, Gastinger I, Lippert H. Hospital caseload and the results achieved in patients with rectal cancer. Br J Surg 2001;88:1397–402.
73. Porter GA, Soskolne CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. Ann Surg 1998;227:157–67.
74. Begg CB, Riedel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. N Engl J Med 2002;346:1138–44.
75. Hosenpud JD, Breen TJ, Edwards EB, Daily OP, Hunsicker LG. The effect of transplant center volume on cardiac transplant outcome. A report of the United Network for Organ Sharing Scientific Registry. JAMA 1994;271:1844–9.
76. Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. Pediatrics 1995;95:323–30.
77. Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. Pediatrics 1998; 101:963–9.
78. Tu JV, Austin PC, Chan BT. Relationship between annual volume of patients treated by admitting physician and mortality after acute myocardial infarction. JAMA 2001;285: 3116–22.
79. Thiemann DR, Coresh J, Oetgen WJ, Powe NR. The association between hospital volume and survival after acute myocardial infarction in elderly patients. N Engl J Med 1999;340:1640–8.
80. Jollis JG, Peterson ED, Nelson CL, et al. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. Circulation 1997;95: 2485–91.
81. Malenka DJ, McGrath PD, Wennberg DE, et al. The relationship between operator volume and outcomes after percutaneous coronary interventions in high volume hospitals in 1994–1996; the northern New England experience. Northern New England Cardiovascular Disease Study Group. J Am Coll Cardiol 1999;34:1471–80.
82. Klein LW, Schaer GL, Calvin JE, et al. Does low individual operator coronary interventional procedural volume correlate with worse institutional procedural outcome? J Am Coll Cardiol 1997;30:870–7.
83. Ho V. Evolution of the volume-outcome relation for hospitals performing coronary angioplasty. Circulation 2000;101: 1806–11.
84. Showstack JA, Rosenfeld KE, Garnick DW, Luft HS, Schaffarzick RW, Fowles J. Association of volume with outcome of coronary artery bypass graft surgery. Scheduled vs nonscheduled operations. JAMA 1987;257:785–9.
85. Shroyer AL, Marshall G, Warner BA, et al. No continuous relationship between Veterans Affairs hospital coronary artery bypass grafting surgical volume and operative mortality. Ann Thorac Surg 1996;61:17–20.
86. Clark RE. Outcome as a function of annual coronary artery bypass graft volume. The Ad Hoc Committee on Cardiac Surgery Credentialing of The Society of Thoracic Surgeons. Ann Thorac Surg 1996;61:21–6.
87. Nallamothu BK, Saint S, Ramsey SD, Hofer TP, Vijan S, Eagle KA. The role of hospital volume in coronary artery bypass grafting: is more always better? J Am Coll Cardiol 2001;38:1923–30.
88. Luft HS. Better for whom? Policy implications of acting on the relation between volume and outcome in coronary artery bypass grafting. J Am Coll Cardiol 2001;38:1931–3.
89. Green J, Wintfeld N. Report cards on cardiac surgeons – assessing New York state's approach. N Engl J Med 1995; 332:1229–33.
90. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829–36.
91. 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.
92. Khuri SF. Invited commentary: surgeons, not General Motors, should set standards for surgical care. Surgery 2001; 130:429–31.
93. Sowden AJ, Deeks JJ, Sheldon TA. Volume and outcome in coronary artery bypass graft surgery: true association or artefact? BMJ 1995;311:151–5.
94. 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;185:315–27.
95. Farley DE, Ozminkowski RJ. Volume-outcome relationships and in-hospital mortality: the effect of changes in volume over time. Med Care 1992;30:77–94.
96. Bronskill SE, Normand SL, Beth LM, Rosenheck RA. Longitudinal profiles of health care providers. Stat Med 2002; 21:1067–88.
97. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 1994;86:829–35.
98. Durazo-Arvizu R, McGee D, Li Z, Cooper R. Establishing the nadir of the body mass index-mortality relationship: a case study. J Am Stat Assoc 1997;92:1312–9.
99. Efron B, Morris C. Data analysis using Stein's estimator and its generalizations. J Am Stat Assoc 1975;70:311–9.
100. McNeil BJ, Pedersen SH, Gatsonis C. Current issues in profiling quality of care. Inquiry 1992;29:298–307.

101. Normand S-LT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. J Am Stat Assoc 1997;92:803–14.
102. Grunkemeier GL, Zerr KJ, Jin R. Cardiac surgery report cards: making the grade. Ann Thorac Surg 2001;72:1845–8.
103. Gatsonis C, Normand SL, Liu C, Morris C. Geographic variation of procedure utilization. A hierarchical model approach. Med Care 1993;31:YS54–9.
104. Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? Health Serv Res 1987;22:157–82.
105. Pronovost PJ, Jenckes MW, Dorman T, et al. Organizational characteristics of intensive care units related to outcomes of abdominal aortic surgery. JAMA 1999;281:1310–7.
106. Daley J. Invited commentary: quality of care and the volume-outcome relationship–what's next for surgery? Surgery 2002;131:16–8.
107. Finlayson SR, Birkmeyer JD, Tosteson AN, Nease RF Jr. Patient preferences for location of care: implications for regionalization. Med Care 1999;37:204–9.
108. Eagle KA, Guyton RA, Davidoff R, et al. ACC/AHA Guidelines for Coronary Artery Bypass Graft Surgery. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery). American College of Cardiology/American Heart Association. J Am Coll Cardiol 1999;34:1262–347.
109. Advisory Council for Cardiothoracic Surgery. American College of Surgeons. Guidelines for standards in cardiac surgery. Bull Am Coll Surg 1997;82:27–9.