Case Volume as a Predictor of Inpatient Mortality After Esophagectomy

Michael Rodgers, MD; Blair A. Jobe, MD; Robert W. O'Rourke, MD; Brett Sheppard, MD; Brian Diggs, PhD; John G. Hunter, MD

Hypothesis: Volume criteria are poor predictors of inpatient mortality after esophagectomy. Because many factors influence mortality for complex procedures, this study was designed to quantify such factors and analyze the volume-outcome relationship for esophagectomy.

Design: Retrospective review of the Nationwide Inpatient Sample database for esophagectomies. We performed multivariate analysis to identify patient and institution risk factors for death and, by using all reported volume thresholds, calculated the probability of choosing a provider with a low mortality.

Patients and Setting: Patients undergoing esophagectomy between January 1, 1988, and December 31, 2000, included in the Nationwide Inpatient Sample database.

Main Outcome Measure: Inpatient mortality.

Results: We identified 8075 cases of esophagectomy; 3243 had complete data sets. The national average mortality rate was 11.4%. Independent risk factors for mortality included comorbidity, age (>65 years), female sex, race, and surgeon volume. Choosing a surgeon or hospital on the basis of a particular volume threshold had a modest influence on the probability of that provider having a low mortality. A low-volume hospital (defined by the Leapfrog Group criterion as <13 cases per year) had a probability of 61% of having a mortality of less than 10%, whereas a high-volume hospital had a probability of 68%.

Conclusions: Patient factors have a greater influence on inpatient mortality than case volume does. Although there is generally an inverse relationship between case volume and mortality, there is wide scatter between individual surgeons and hospitals, with a complex volume-outcome relationship. Using volume criteria alone to choose a provider may in some instances increase the risk of mortality.

Arch Surg. 2007;142(9):829-839

After the seminal 1979 article by Luft et al, numerous studies have described the relationship between increasing case volume and improved outcome. Certain operative procedures are more affected than others by volume, including esophagectomy. Recent studies based on 2 of the largest administrative databases in the United States have shown that high-volume hospitals and surgeons performing esophagectomy have inpatient mortality rates that are half those of low-volume centers. A natural consequence of this has been policy aimed at directing patients to high-volume centers. This has been most systematically applied by the Leapfrog Group, which consists of more than 150 health care purchasers and currently advocates that esophagectomy be performed only at institutions with an annual caseload of at least 13. Attempts have been made to estimate the number of lives that would be saved by these policies. Birkmeyer et al estimated that 168 lives a year in the United States would be saved if the volume criteria were applied to esophagectomy.

A potential problem of using volume criteria is that it involves applying an average result to an individual provider. Although it might intuitively be expected that a given provider with a higher volume of cases will have better outcomes, this is not necessarily true. The estimate of lives saved may very much depend on which high-volume institution the patients were referred to.

The present study was designed to investigate the effect on mortality of all available patient and institutional factors and to explore the relationship between volume criteria and outcome in more detail.
METHODS

The Nationwide Inpatient Sample (NIS) database, a part of the Healthcare Cost and Utilization Project by the Agency for Healthcare Research and Quality, is an administrative database representing a 20% sample of US hospitals. The sample is weighted to include representative numbers of hospitals defined by region (West, Midwest, South, and Northeast), location (rural vs urban), teaching status, ownership status (public vs private), and number of beds. Federal hospitals are not included.

The NIS database was queried for cases with International Classification of Diseases, Ninth Revision, Clinical Modification procedural codes indicating partial or total esophagectomy (codes 42.40, 42.41, and 42.42). Analysis was limited to adults older than 18 years. We reviewed the diagnostic codes, and patients without a primary esophageal diagnosis were excluded to account for cases in which a partial esophagectomy was performed as an adjunct to another procedure such as pharyngectomy, laryngectomy, pneumonectomy, or total gastrectomy.

Descriptive statistics were used for reporting patient and institution variables. Available patient variables included age, race, sex, income, emergent vs elective surgery, and benign vs malignant pathologic findings. Secondary diagnoses that represent comorbidity and postoperative complications were also available. For this analysis, age was categorized as younger than 65 years vs 65 years or older; race was categorized as black vs nonblack; and years of operation were categorized as 1988 to 1996 vs 1997 to 2000. Income was assessed at the zip code level and grouped into categories of less than $25,000, $25,000 to $35,000, and more than $35,000. Available institution factors included geographic location (West, Midwest, South, or Northeast), hospital type (urban teaching, urban nonteaching, or rural), and surgeon and hospital case volumes.

We calculated inpatient mortality rates for esophagectomy and used logistic regression to identify which variables acted as independent risk factors. Data for the multivariate analysis were taken only from cases in which all variables were reported. In this analysis, surgeon and hospital volume were treated as continuous variables. We also corrected for the year of procedure.

Secondary diagnoses are coded in the NIS database and may represent a preoperative comorbidity or postoperative complications. Therefore, we limited our analysis to chronic conditions, including obesity, diabetes, chronic pulmonary disease, hypertension, valvular disease, and peripheral vascular disease.

We also assessed the overall effect of volume on mortality rate (MR) by using the volume categories set by the Leapfrog Group (≥ 13 vs < 13 cases per year) and by Finlayson et al (< 4, 4-9, and > 9 cases per year). Surgeon volumes were assessed similarly using the criteria set by Birkmeyer et al ( < 2, 2-6, and > 6 cases per year). Differences in mortality between these volume categories were calculated while correcting for all other patient and institutional variables.

For analysis of the utility of volume criteria to predict mortality, volume threshold (VT) was defined as an annual case volume used to divide providers (surgeons and hospitals) into high- and low-volume groups. A mortality threshold (MT) was defined as an inpatient mortality rate used to divide providers into high- and low-mortality groups. For any given VT and MT, the probability of picking a low-MR provider can be calculated. To simplify the analysis and maintain clinical relevance, 3 MTs were set at 0.5, 1.0, and 1.5.

Figure 1. Odds ratios and 95% confidence intervals for patient factors and inpatient mortality. Odds are calculated using univariate and multivariate analyses.
chosen based around the average MR in the NIS database, ie, 5%, 10%, and 13%.

RESULTS

In the 13 years from January 1, 1988, through December 31, 2000, there were 8785 cases with International Classification of Diseases, Ninth Revision, Clinical Modification codes of 42.40, 42.41, or 42.42. Patients were younger than 18 years in 181 cases, and, in 2 cases, the age was missing. In 527 cases, we found diagnostic codes for non-esophageal disease. The remaining 8075 cases were available for further analysis. Of these, 3607 were missing surgeon-level data, 2500 were missing race, 339 were missing income, 7 were missing the outcome variable of in-hospital mortality, and 2 were missing hospital location and teaching status. No cases were missing sex or region. By the previous selection criteria, none of these cases were missing age or malignancy information, and the comorbidity identification procedure does not create any missing data. A total of 3243 cases had all variables of interest and were considered in the multivariate analysis.

When the sample is weighted to allow estimation of national results, 46 562 esophagectomies were performed in the 13-year study period, with an annual average of 3582. The inpatient MR averaged 11.4%, with a high of 13.95% in 1988 and a low of 8.41% in 1999, but with no significant trend over time. In the last 5 years of the study period, the MR averaged 10.2%. Of the sample, 50.0% were older than 65 years, 75.8% were men, and 8.1% were black. For income, 46.3% of cases were in the highest bracket (> $35,000); 31.6%, in the middle bracket ($25,000 - $35,000); and 22.1%, in the lowest bracket (< $25,000). Of the esophagectomies performed, 83.4% were for a malignant neoplasm.

Patient variables significantly increasing the odds for inpatient death were sex, age, and race (Figure 1). Women had a 1.5-fold increase in the odds of death, whereas patients older than 65 years and black patients had double the odds of death.

The additional diagnoses were assessed for their effect on mortality (Figure 2). Diagnoses that most likely represent preoperative comorbidity, such as chronic pulmonary disease, valvular heart disease, diabetes mellitus, and obesity, did not achieve significance as risk factors for inpatient mortality. The exceptions were peripheral vascular disease and hypertension. Hypertension appears protective, but this may be owing to its high prevalence, causing it to be coded only in the absence of other, more acute conditions.12

Geographic region did not have a significant effect on mortality (Figure 3). Urban hospitals were no better than rural hospitals. Although teaching status appeared to confer benefit in the univariate analysis, this lost significance once hospital volume was included. Hospital volume was highly significant in results of multivariate analysis.
analysis when surgeon volume was ignored. However, hospital volume failed to achieve significance when surgeon volume was added to the analysis (Figure 4). By contrast, the effect of surgeon volume remains, even with the addition of hospital volume (odds ratio per additional case, 0.90).

The analysis of case volume as a categorical variable is shown in Figure 5 and Figure 6. The VTs for hospitals are those determined by Finlayson et al\(^8\) and the Leapfrog Group. For surgeons, we used the VTs of Birkmeyer et al.\(^9\) The plot emphasizes the considerable scatter within each group. There is a large number of low-volume providers with a low MR and a significant number of high-volume providers with a high MR. The adjusted average MR for high-volume hospitals (>9 cases per year) was 11.55%; for medium-volume hospitals (5-9 cases per year), 7.78%; and for low-volume hospitals (1-4 cases per year), 11.37%. Therefore, when individual surgeon case volume was controlled for, we found no difference attributable to hospital case volume. The average MR for high-volume surgeons (>6 cases per year) was 9.25%; for medium-volume surgeons (2-6 cases per year), 7.46%; and for low-volume surgeons (<2 cases per year), 12.75%. Although this shows an overall improvement in MR with increased surgeon volume, it is not as great as previously reported.

The Table shows the probability of a provider in each volume category having an MR that is less than the threshold. Despite the mean MR being lower for the high-volume group, the probability of picking an individual surgeon with an MR below a given threshold is actually lower for the high-volume group (Table).

A similar analysis of all VTs is shown in Figure 7. The probability of choosing a provider, from among the high-volume providers, with a low MR for all VTs is plotted. Mortality thresholds of 5%, 10%, and 15% are used. The points corresponding to the lowest VT represent the probability of choosing a provider with a low MR from among all providers. Only when the curves rise above this initial value does using volume to aid provider se-
lection confer a benefit. The degree to which the curve rises above this initial point is a measure of the degree of benefit.

**COMMENT**

There are limitations to using administrative databases for clinical data. Data entry is based on financial rather than clinical considerations, and this carries the risk of systematic error when interpreted clinically. An example of this is the apparent positive relationship between hypertension and mortality. This relationship was apparent for a number of diseases when the comorbidity variables were first analyzed by Elixhauser et al. It has been explained as a systematic bias in coding. Patients with more severe comorbidities will have these comorbidities preferentially coded and hypertension may be left out, whereas otherwise healthy patients will be more likely to have this very common condition coded. There may be other systematic coding biases that go unrecognized.

Another limitation is the lack of important clinical factors such as cancer stage, the use of neoadjuvant and adjuvant therapy, and other clinical data that might affect outcome. Moreover, recorded outcomes are limited to inpatient mortality and length of hospital stay. Complications are not coded separately and cannot be reliably determined from the database. Nonetheless, in the absence of a nationwide clinical database, these administrative databases represent our only way of assessing national demographics, trends, and outcomes for procedures.

The NIS is an admission-based database, and patient transfers are counted as separate admissions. If a patient was moved to another institution after the esophagectomy and died, that death would not be captured. This suggests that the average mortality of 10.2% in the past 5 years may be an underestimate. It is also likely that some low-volume centers would have higher mortality rates if transferred patients were taken into account. We are unable to correct for this.

Surgeons practicing at multiple institutions may have their volume underestimated because physician identifiers are not guaranteed to be the same in different hospitals. In addition, there is no guarantee that both hospitals would appear in the sample, so only the cases from the sampled hospitals would contribute to the surgeon volume.

The series from Queen Mary Hospital, Hong Kong, M. D. Anderson, Houston, Texas, and University of Munich, Munich, Germany, point out that their MRs are continuing to improve and in the last 5 years of data collection were 1.1%, 3%, and 6%, respectively. This is considerably lower than the national average for the last 5 years of our study period of 10.2%. This finding highlights the fact that centers of excellence can achieve outstanding results, but it is also true that many smaller centers have similarly outstanding results.
It is unclear why women have a higher MR than men. Similarly, the relationship with race is as yet unexplained. The finding that black patients have a higher mortality than nonblack patients, even accounting for all other factors, requires further investigation. It may be related to the stage or the cell type of the cancer, but we cannot answer that question from these data. That age was found to be an independent risk factor seems intuitively obvious and is borne out by the recent prospective series of 421 esophageal resections from Hong Kong,13 in which the MR for patients aged 76 to 80 years was 24% and for those aged 81 to 85 years was 33%, whereas the overall MR was 4.8%.

Contrary to previous findings, our results indicated that including surgeon volume eliminated the independent effect of hospital volume on MR. Birkmeyer et al9 found that only 46% of the hospital volume effect was accounted for by surgeon volume. However, in that study, the addition of surgeon volume to the logistic regression brought the lower limit of the 95% confidence interval for the odds ratio for hospital volume to 1.02, making it very much on the borderline for statistical significance. With our data set, this statistical significance was eliminated altogether. In any case, it seems clear in the current analysis that surgeon volume is a more important factor than hospital volume.

This study highlights the difficulty of using average population results to predict outcomes in individuals. We found the same qualitative result, that high volume is in general associated with a lower average MR, at least for surgeon volume. However, the application of that result to predict outcome in particular providers is flawed. It ignores the wide scatter of results that are found between providers at all case volumes. This is highlighted by the fact that one hospital with a caseload of more than 13 per year had an MR of 25% and one surgeon with a case load of more than 6 per year had an MR of 40%. Choosing those particular providers on the basis of volume might well be a mistake.
Figure 7. Probabilities of picking a hospital (A) or a surgeon (B) with a mortality rate (MR) below the threshold when selecting from those hospitals or surgeons, respectively, with a volume greater than the volume threshold.
The relationship between volume and outcome is complex and certainly not linear. Moreover, Figure 7 demonstrates that the utility of VTs to predict outcome is dependent on the MT. This finding suggests that volume criteria should be defined in relation to agreed outcome thresholds. The reasons for the complexity of this relationship are unclear. Some of the good mortality results in the very-low-volume groups probably reflect the relative ease of achieving 100% survival in 1 or 2 patients. However, even if we excluded providers performing an average of less than 1 esophagectomy a year, there was no appreciable difference in the results. This wide variation in individual results has also been described in coronary bypass surgery (Welke et al16), in which situation the use of volume criteria to pick a provider is described as “slightly better than a coin flip.”

However, we believe that the present study does not negate previous findings that high case volume is associated with better results on average. We do not see the results of our study as an argument against centralization of complex and expensive services. It is still most likely that concentration of such services into centers of excellence will be beneficial. However, establishing what constitutes a center of excellence is not simply a matter of volume. In a particular instance, it is very possible that using volume criteria alone could result in a patient being transferred from a unit with excellent outcomes to one with inferior results. The problem of using indicators such as volume to predict outcome could be solved if the outcome data were available for all institutions and surgeons. This would be a difficult and controversial undertaking.

An alternative strategy to volume criteria might include national benchmarks for outcome. Institutions could nominate themselves as places where such procedures are performed regularly. It would then be up to the institution to demonstrate that they meet set benchmarks. If they fail to do so, then the purchasing group would send its patients elsewhere. One possible method for establishing benchmarks would be to use the national result for a given outcome averaged over the past 5 years. For example, with esophagectomy this would mean that the benchmark for annual inpatient mortality would be less than 10%. A reasonable time frame for an institution to establish that it is meeting this benchmark might be 5 years. Choosing the average mortality as the benchmark puts pressure on approximately half of the institutions performing this operation to improve their performance or stop doing the procedure. Making the institution responsible for showing that they are meeting such a benchmark means they must record and review their outcomes. This would give a real incentive to maintain a genuine audit and quality control program.

Such a system could have other potential benefits. Currently, if a surgeon encounters a high-risk patient, there is a financial disincentive to refer that patient to a more experienced institution. Moreover, if volume is the criteria for a payer group to refer patients, this disincentive is even stronger. However, with a benchmark system, the high-risk patient would be more likely to be referred, so that the potential mortality does not cause the institution to exceed the benchmark. In fact, the fewer cases that institution treats, the more likely the institution will refer the high-risk patient because the potential effect on mortality rates will be greater.

Another potential benefit to patients would be the scrutiny placed on the individual surgeon’s results. Because of the possibility that the institution could lose its referral base if the benchmark is exceeded, there may well be a shift within surgical departments to concentrate cases in the hands of the surgeon with the best outcomes.

A benchmark-based system is intrinsically less confrontational and potentially more fair than a volume-based system. The latter clearly excludes some institutions with excellent results. A benchmark-based system simply sets clear guidelines and allows institutions and surgeons to find their own means to achieve them. In the medium term, it would also reassure patients that the institution they were going to had satisfactory and verified outcomes for that procedure.

Inpatient mortality for esophagectomy is more strongly associated with patient variables than institution variables. The association between hospital and surgeon case volume and average annual inpatient mortality has been confirmed, although surgeon case volume is the more important factor and eliminates hospital volume in multivariate analysis. However, case volume criteria are poor indicators of inpatient mortality rates for a particular hospital or surgeon. The adoption of a national benchmark system for outcomes of complex surgical procedures may be an alternative to achieve the goals of improved quality of health care and improved patient outcomes.

Accepted for Publication: April 17, 2007.
Correspondence: Blair A. Jobe, MD, Department of Surgery, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd, L-223A, Portland, OR 97239-3098 (jobe@ohsu.edu).
Author Contributions: Study concept and design: Rodgers and Diggs. Acquisition of data: Jobe, O’Rourke, and Hunter. Analysis and interpretation of data: Rodgers, Sheppard, and Diggs. Drafting of the manuscript: Rodgers and Hunter. Critical revision of the manuscript for important intellectual content: Rodgers, Jobe, O’Rourke, Sheppard, Diggs, and Hunter. Statistical analysis: Diggs. Obtained funding: Jobe. Administrative, technical, and material support: Jobe, O’Rourke, and Hunter. Study supervision: Jobe, O’Rourke, Sheppard, and Hunter.
Financial Disclosure: None reported.
Funding/Support: This study was supported in part by grant 5K23 DK066163-05 from the National Institutes of Health (Dr Jobe).
Previous Presentations: This paper was presented at the 78th Annual Meeting of the Pacific Coast Surgical Association; February 20, 2007; Kohala Coast, Hawaii; and is published after peer review and revision. The discussions that follow this article are based on the originally submitted manuscript and not the revised manuscript.
adjusted mortality among hospitals in the same volume categories or the fact that there is considerable overlap in mortality rates between hospitals in different volume categories.

The authors correctly point out that using volume as a proxy for quality is not 100% accurate and very reasonably suggest using outcome data. However, the “reasonable” 3-year time frame they suggest to allow institutions to establish that they are meeting the benchmark isn’t good enough. Dr Birkmeyer estimates that 168 lives a year could be saved by centralizing esophagectomies. Until outcome data are available, I feel we must centralize high-risk procedures like pancreatectomy and esophagectomy. If a low-volume hospital can demonstrate acceptable mortality rates over several years, it should be listed as a center of excellence, and if a high-volume center reports poor numbers, it should be on probation and an immediate plan of action initiated. Failure to demonstrate improvement should result in a failure to be endorsed as a center of excellence.

Dr Birkmeyer’s data indicate that the mortality rates for esophagectomy are related to both hospital and surgeon volume. A low-volume hospital may have a single experienced surgeon performing esophagectomy with excellent results, whereas a larger system may have some low-volume surgeons with poor results that are submerged into acceptable institutional numbers. Should surgeons have their own quality data compiled and available for comparison by Leapfrog or similar groups? Finally, what do the authors feel we should do with recent data from Dr Birkmeyer that volume not only affects short-term mortality but also 5-year survival in esophagectomy? Survival dramatically doubles from 17% at low-volume centers to 34% at high-volume centers. How many lives a year might we save by centralization while we await quality data? Dr Hunter: Dr Standage, thanks for helping to focus the thesis of this paper, which is not to say that volume isn’t an important marker for quality but that volume is not the best marker to predict quality. If patients use esophagectomy volume alone as the criterion for picking their surgeon or hospital, they will be marginally better off than a coin flip. This is true because there are high-volume hospitals with poor outcomes and low-volume hospitals with superb outcomes. Bottom line: you have to look beyond esophagectomy volume alone.

Dr Standage, you inquire about the need to wait 5 years, collecting data prospectively before determining whether a center meets mortality benchmark criteria. I don’t think this is necessary, as we can use retrospective data from state databases, NIS, University HealthSystem Consortium, National Surgical Quality Improvement Program (NSQIP), or a number of other data sources. The data must be sufficiently robust to capture those who die after hospital transfer or within the first few weeks after hospital discharge. To be inclusive, especially at the outset, we would suggest that we set a reasonable mortality rate of 10%. Half the hospitals performing esophagectomy will fall within this range. Then one can continually increase quality through collaboration (as did the Northern New England Cardiovascular group) to the point where you actually may be able to achieve mortality rates of 7%, 6%, and 5% in the benchmarked hospitals.

You talked a little bit about credentialing surgeons. I think this segues nicely into the next paper and some of the comments I will have on that. While surgeon volume was a stronger predictor of outcome than hospital volume in this analysis, as we develop multidisciplinary teams of care, the outcomes that we report in the next decade may better reflect the team and the environment than it will the individual surgeon. Therefore, I believe it is the responsibility of the hospital, if it wishes to maintain a “benchmark” mortality rate, to winnow out the surgeons within the institution who can’t perform at that benchmark rather than allowing payers to make this determination. Again, I think the next paper may point out some of the virtues of that.

REFERENCES


DISCUSSION

Blayne Standage, MD, Portland, Oregon: Can volume be used as a proxy for quality and a means of selecting hospitals or surgeons in high-risk surgery? Although for a common operation like coronary artery bypass grafting, where mortality differences are small, it has been said that using volume alone to select is little better than a coin flip, surgical oncology is different. Some operations, notably pancreatic resection and esophagectomy, have large differences in volume-related outcomes and may warrant centralization. Birkmeyer’s NIS data are dramatic. At a low-volume center, you have a 15% chance of dying after your esophagectomy vs. 6.5% for high-volume centers. Finlayson and Brennan, using linked Surveillance, Epidemiology, and End Results–Medicare data, showed an even greater difference, 17.3% vs 3.4%.

Using an administrative database has disadvantages, as pointed out in Dr Johes’s paper. We must use in-hospital mortality as the end point, and we can’t stratify for risk, yet studies suggest that using either an administrative or a clinical database gives similar results. While it is likely that higher volumes do lead to better outcomes, other possibilities exist. Better outcomes may result in improved referrals and thus higher volume. Mortality differences appear to be quite high, but raw numbers do not demonstrate the wide variation in the risk—
I think that your last comment reflected on a recent paper, which showed that the likelihood of a 1-year survival after a cancer operation is better at a high-volume center. Agreed, but as a single factor for picking your surgeon or your institution, procedure volume does not provide enough information. Thank you very much, Dr. Standage.

Sherry Wren, MD, Stanford, California: Thank you for this interesting presentation, because those of us who are low-volume and low-mortality surgeons love a paper that justifies our still doing this operation. I think we have all begun to realize that the whole process of care influences the patient’s outcome, which is illustrated by your data. It’s not just the surgeon volume or the hospital volume as a single factor. Can you use this database to look at process and systems factors such as whether these are American College of Surgeons oncology-certified hospitals, level I trauma centers, or what the volume of cardiac cases in a year is? I think esophageal mortality has to also do with resources present to take care of these complex patients such as intensive care unit (ICU)—experienced physicians, information resources, and other hospital resources, which are difficult to represent in mortality-to-volume single-factor analysis. Are you expanding this study to look at some of the other variables?

Dr. Hunter: What you are pointing out is that it’s not just the outcome data that should be measured. Most quality measures, such as those used by Leapfrog, assess structure and process issues (Does the hospital have 24-hour intensivists? Are antibiotics administered on time? etc). I think we would agree that we need structure, process, and outcome data to reasonably assess quality. Today, no one repository of data yet provides us all of those things. The NIS data set is a purely administrative data set and looks just at hospital and surgeon-related events. It doesn’t provide process measures, whereas there are data sets such as the Surgical Care Improvement Project (Centers for Medicare and Medicaid Services) that look only at process measures. In the future, NSQIP may be able to couple structure, process, and outcome measures to provide the best assessment of quality at a particular institution.

Dr. Wren: Our group is currently starting a project using the Veterans Affairs NSQIP database, indexing it to other hospital variables such as cardiac patient flow, ICU admissions, for esophagectomy, Whipple resections, and other high-risk surgical procedures.

Dr. Hunter: Thank you. We will look forward to it.

Ralph W. Aye, MD, Seattle, Washington: Very nice paper and an important rebuttal to the concept of using volume alone as a measure of quality for surgeons. We all know there are surgeons who do everything well in a variety of areas, even though their volume in a particular area may not be so high. We have actually seen surgeons at our hospital lose privileges in a specialty area—vascular, for example—for not doing a set number of cases, even though they were still doing a good job. However, the problem I have with this paper is that I don’t have the sophistication to critique your methodology in detail, but I’m skeptical of the results. I would just ask you to explain why you think your conclusion is so at variance with the bulk of the literature. There are at least 4 esophagectomy-specific papers that show significant volume-related differences in mortality and many other complex procedures that have been shown to have similar outcomes related to volume, both for immediate mortality and morbidity issues, as well as for long-term survival for cancer patients.

Dr. Hunter: The first part of this paper confirmed that, using the NIS database, the volume-outcome relationship could be demonstrated and was more strongly associated with surgeon volume than hospital volume. The novel observation of this study is that when one takes a “patient-centered” approach to using these data, the relationship is not nearly as valuable. When you ask the question “Is my risk of esophagectomy higher at hospital A with surgeon B than it is at hospital C with surgeon D?” procedure volume is not of great assistance in improving your odds of survival. Nonetheless, there is a benefit accrued by picking a high-volume surgeon and hospital, but the benefit is not large. It is much better to try to determine the real, observed track record of the surgeon and the institution. You have to look beyond the volume data. You can’t use volume data alone as an effective surrogate for quality data.

Robert Cameron, MD, Los Angeles, California: I agree that volume data are not certainly the best predictors necessarily in all circumstances. In fact, I operate at a low-volume center and a high-volume center, and the results are pretty much equivalent. My question regards using mortality alone as the best predictor for outcomes from esophagectomy. One can have somebody who gets operated on by a surgeon and spends up to 3 weeks in the ICU, gets out debilitated, can’t swallow, and ends up going to a nursing home, and that person survives, but that outcome is far different than that of somebody who has been operated on, leaves the hospital in 5 days, swallows really well, and does really well from the oncology standpoint. So my question is, should there be a more complex model to figure out what is an adequate outcome, because I know in our center we can have different surgeons operating on patients and the morbidity is incredibly different but the mortality outcome is the same?

Dr. Hunter: You are absolutely right. This is a 30,000-foot view and doesn’t get into some of the important questions that you might want answered. Using any administrative data set, there are significant limitations. One of the things I showed you was that rural hospitals and urban hospitals had the same mortality rates for esophagectomy. Do we really believe that? Probably not. So what happened? Well, it may very well be that when the patients got sick in the rural hospitals, they were transferred to the urban hospital. If the patient dies, that mortality doesn’t get recorded for the urban hospital because the esophagectomy wasn’t done there, but it also doesn’t get recorded for the rural hospital that should have had it tagged to them. So there are confounders when you look at these types of data. You have to be aware of that.

Carlos A. Pellegrini, MD, Seattle: An ethical question now. Should it be standard of care for surgeons to disclose to their patients the number of operations that they have done for a given disease? The data are in general pointing to the fact that complex operations have a different outcome in the hands of people who have done more than in people who have done fewer. Should that become a standard of care?

Dr. Hunter: People are always asking us about volume. I think we have to be honest about our volume. But I actually try to take that one step further. The procedure I have done the most of is Nissen fundoplication. Not only do I offer the number of procedures that I have performed, but I feel strongly that the data on failure, complication, and reoperation need to be disclosed. You don’t want to tell them necessarily about mortality with Nissen because the mortality rate is diminishingly low, but you do have to talk about your mortality rate with esophagectomy. What is the mortality rate of esophagectomy in your hands? Also germane is the mortality rate of esophagectomy in your center.

Ronald G. Latimer, MD, Santa Barbara, California: I was just taking a note on Dr. Pellegrini’s point. I don’t do the big, complicated operations. I have limited mine to breast surgery and thyroid surgery at this point, but my patients are asking me when they come through the door how many of these have you done and how many mastectomies have you done? So that’s a frequent question every day, which I have to respond to. For the thyroid patients, I inform them that I have injured a recurrent laryngeal nerve. To me this is informed consent, as Dr
Michael Hart presented in his great Presidential Address. The reason I rise is not necessarily to discuss the paper but to comment about a fear that I see coming. I go back historically to when the single-payer system came in for renal disease and at that time the federal government decided to centralize renal transplant programs. They closed little programs to centralize this care because of the concept that centralized high-volume care is better care, although the data did not support that contention. The actual reason they closed the small centers was cost. In some small centers, we were actually able to prove that the costs were less. Nonetheless, they closed them. My fear is that we are going to centralize on the basis of volume and hospital costs. In Europe, people go to centralized hospitals. There are hospitals that only do gastric surgery or esophageal surgery, and are we going to get to that point where we are going to be told to transport patients to specific centers for certain procedures? That is how the future may develop, and to me this is a concern.

Dr Hunter: Again, I hope that this paper helps dissuade policy makers from relying on volume criteria only to make those determinations. I think quality criteria as defined by structure, process, and outcome should drive these decisions. But hospitals need to be measured, and it may well be that some small, low-volume hospitals might provide outstanding esophagectomy outcomes. I see no reason why you should make someone from Santa Barbara travel to Los Angeles and get a product that actually may, when you measure it, not be as good as they would receive by staying in Santa Barbara. I’m not picking on any center in Los Angeles because they are great, but that may be the truth.

Thomas R. Russell, MD, Chicago, Illinois: My question is sort of a follow-up to Dr Latimer’s question. Traditionally, hospitals in this country have tried to do everything at all stages of a patient’s life, from womb to tomb. A good competitive hospital had to do everything. Now we are entering an era of transparency with respect to outcomes reporting and, of course, cost reporting, adding up to value. Do you think that information like you have given us this morning is going to make hospitals really look at the book of business they are doing and opt out of providing certain services? I noticed you had a couple of red bullets. For example, your hospital may opt out of doing pediatrics and send patients needing those services to another hospital in Portland that has a really strong record in that specialty. Meanwhile, that hospital would opt out of doing, for example, these high-end procedures that vary with respect to outcomes. Not only are they high risk, but the value sometimes is questionable. So the question relates to hospitals not having to do everything and making conscious decisions about the work they want to do.

Dr Hunter: That’s a great question, and I will forward your comments to our hospital administrator because that would open up many more beds for surgery. Dr Russell, you are always one step ahead of everybody else. That’s exactly what we ought to be doing within communities, and, in fact, we are working with the Legacy system to do just this. We have 2 children’s hospitals in Portland and we are trying to decide who does what where, especially as we link our pediatric cardiac surgery programs. The opportunity here is to improve the quality of our outcomes as well as decreasing cost by centralizing certain services. It takes a lot of forward-thinking health administrators to actually do some of that. I am not sure many communities have that. We are lucky.

William P. Schecter, MD, San Francisco, California: Dr Hunter, where do the underserved and disenfranchised fit into this equation? Everybody talks about quality, but if I want to send one of my uninsured patients to a “high-quality” center and they don’t have money, they may say thanks but no thanks. So we do everything. How does that problem get resolved when there are over 40 million people in this country without health insurance?

Dr Hunter: That is a great question. I think that some cities have public health systems and others don’t. In our city, I can tell you what happens. Forty percent of our inpatients at Oregon Health & Science University are either underserved or uninsured. We serve a large portion of the state “safety net.” Yet, we are the highest-volume center for esophagectomy in the state, with high-quality outcomes for a variety of procedures, including esophagectomy. I would think that cities that do have public hospitals might be challenged to develop that sort of expertise within their system. If certain public hospitals can’t provide the quality, then these hospitals should use this as an incentive to improve, or they should make a “deal” with the center capable of delivering high-quality specialty care. We may have 2 tiers of hospitals, but we cannot allow ourselves to accept 2 standards of care and become complacent about disparate mortality rates offered to patients without the means to fly to a clinic that specializes only in their condition.

Financial Disclosure: None reported.