Esophagectomy volume threshold as a criterion for centers of excellence: Causation or cause, strategy or strategem?

Thomas W. Rice, MD, and Eugene H. Blackstone, MD

PROLOGUE
Outcomes ranking, number of stars received, and designation of excellence by a host of watchdog organizations, government agencies, insurers, and national societies have captivated and captured every hospital, program, and department chair. The cause is to promote higher-quality, yet lower-cost, medical care (value). The pivotal target of this cause was hospital mortality after coronary artery bypass grafting—a high-cost, commonly performed, rather uniform, routine operation. Esophagectomy is, in contrast, uncommon, complex, and performed for a spectrum of esophageal diseases, both benign and malignant. Groups such as Leapfrog believe that esophagectomy should not be a cottage industry but rather regionalized, much like surgery for congenital heart disease. Their strategy to effect change in surgical practice is to set a threshold (lower bound) for number of operated cases that must be performed each year to qualify for the coveted designation “Center of Excellence.” They promote their strategy as “evidence-based hospital referral.”

It is important to recognize that implicit in this strategy is a direct cause-and-effect relation (causation) between patient outcome and number of cases performed in a hospital or by a surgeon per year, and not merely an association between hospital mortality and volume. The theory behind the strategy is that regionalization will further increase the volume of Centers of Excellence and, thereby, decrease mortality (we have called this neutralization of the effect of low volume). Although this theory may turn out to be fact, and in any event, regionalization may be philosophically compelling for good reasons, Meguid and colleagues in this issue of the Journal of Thoracic and Cardiovascular Surgery question the fundamental evidence for this volume threshold, challenging the theory behind the strategy. Implicit in their discussion is that a volume threshold for esophagectomy as a criterion for a Center of Excellence is a strategem.

Yet, using the same database (although including older data), Rodgers and colleagues seem to arrive at different recommendations. Why? In this editorial, we re-examine and re-evaluate the evidence, starting with the data source, cases analyzed, volume definition, variables examined, outcome assessed, analyses performed, and manner by which differing conclusions and recommendations were reached.

DATA SOURCE
Source of data for this study was the Nationwide Inpatient Sample, a 20% sample of hospital discharges. This administrative database was designed to be representative of hospital characteristics: region in the United States, rural versus urban, teaching versus nonteaching, public versus private, and size. It is part of the Healthcare Cost and Utilization Project and is driven by financial considerations; it is not designed to optimally represent any specific disease or procedure, such as esophagectomy, or to answer clinical questions. Because esophageal surgery is a cottage industry and small programs far outweigh large ones, a consequence of the 20% sample is that no truly large-volume centers performing 80 to 100 esophagectomies per year were captured. Only 4080 esophagectomies were performed at 1506 hospitals in the 6-year study period (1998–2003). The average number of esophagectomies performed per sampled hospital per year was 0.45, median number of esophagectomies per year was 4 among hospitals that did any of these procedures, 75% of these hospitals performed 9 or fewer esophagectomies annually, and the largest annual volume was 34. A practice-specific database, such as The Society of Thoracic Surgeons’ General Thoracic Surgery Database, will be a better source of data for studies like this, because as it progressively penetrates the majority of thoracic surgery programs, it will become increasingly representative of specific thoracic diseases and procedures.

VOLUME DEFINITION
Case volume refers to number of resections performed per year. One question is “by whom?” Meguid and colleagues examined only hospital volume. Rodgers and colleagues examined both hospital and individual surgeon volume and found that surgeon volume was more strongly associated with mortality than hospital volume. One must be careful in reading and interpreting volume–mortality relationships, because different definitions of “surgical volume” may be assessed.

CASES ANALYZED
An esophagectomy is required for inclusion in any study that addresses the volume–mortality relationship for
esophagectomy. Meguid and colleagues\(^3\) included all patients with a diagnosis of esophageal cancer and ICD-9 codes for esophagectomy. Using the same database, despite a study period (1998–2000) that was nearly twice as long, the report by Rodgers and colleagues\(^4\) included a similar number of esophagectomy patients, a result of different inclusion criteria. Rodgers and colleagues\(^4\) were more rigorous in their definition of esophagectomy and excluded patients who did not have esophagectomy for a primary esophageal diagnosis. This eliminated simpler procedures such as partial cervical esophagectomy with jejunal interposition for pharyngeal problems, which potentially have a different risk of mortality than esophagectomy for a primary esophageal diagnosis.

Thus, Meguid and colleagues\(^3\) were specific for diagnosis, but unspecific for procedure; Rogers and colleagues\(^4\) were unspecific for esophageal diagnosis, but specific for procedure. A more rigorous definition of esophagectomy potentially provides a more uniform population for analysis. A population defined by diagnosis may include patients undergoing a procedure that is not "truly" an esophagectomy. Caveat emptor: reader beware.

### VARIABLES EXAMINED

Variables available, how these variables may have been transformed, and which were used in the analysis affect results. Rodgers and colleagues\(^5\) included more patient, socioeconomic, and hospital variables, such as patient income and hospital region, than did Meguid and colleagues.\(^3\) The Rodgers study\(^4\) listed individual comorbidities and found that coding irregularities resulted in a spurious association of hypertension with good outcome. Meguid and colleagues\(^5\) homogenized comorbidities into a single variable, the Charlson Index, that could hide such errors. Some continuous variables were dichotomized, and variables such as indication for esophagectomy, emergency versus elective esophagectomy, and technique and extent of esophagectomy were not available or included in the analysis. Neither study could include important risk factors for early mortality, such as patient condition at operation, pulmonary function, cancer stage, preceding chemotherapy, and so forth, which would be available for risk adjustment in a clinically oriented database.

### OUTCOME ASSESSED

Outcome was in-hospital mortality. Deaths occurring in the postoperative period outside the hospital in which the esophagectomy was performed were not available. Therefore, transfers of terminal postoperative patients to other centers or deaths within 30 days of operation—data necessary to define postoperative mortality—were not included. Studies of early risks in sick patients undergoing major operations for complex diseases have shown consistently that early risk of death extends beyond immediate hospitalization and is considerably underestimated by in-hospital mortality.\(^5,5\)

### ANALYSES PERFORMED

Rodgers and colleagues\(^5\) explicitly describe how they handled missing data; Meguid and colleagues\(^3\) do not. The strategy used by Rodgers and colleagues\(^5\) was to delete cases with missing data, resulting in eliminating five-eighths of the cases! How representative is the remainder? Techniques such as multiple imputation preserve precious data without bias.

Both articles confirm a statistically significant association between low esophagectomy hospital or surgeon volume and increased mortality. Yet neither does the obvious: display graphically a nomogram of the volume–mortality association (with careful attention given to a possible nonlinear relationship), accompanied by confidence limits. Such a nomogram would enable finding evident differences.\(^6\)

Instead of the obvious, Meguid and colleagues\(^3\) adopted a goodness-of-fit analytic strategy, which is removed from the issue being evaluated: the volume–mortality association. They examine sensitivity and specificity for a complete series of volume thresholds. Rodgers and colleagues\(^5\) adopted an innovative analytic strategy, asking what the probability was of finding a low-risk hospital or surgeon. Although one might anticipate the answer would be to find a high-volume program, the answer is to find the lowest volume program! This apparent anomaly is easily explained. There are numerous hospitals and surgeons whose annual case volume is 1; mortality must therefore be either 0% or 100%! Thus, there is a high probability of finding a program with 0% mortality in any given year among these many institutions. The trouble is, the confidence limits are huge, and the chances that zero mortality will be true the following year or for any given future patient are low. Interestingly, confidence limits are not included in either of these papers!

This brings us to interesting Figure 1 in the article by Meguid and colleagues\(^5\): the volume–mortality relation. The lower the case volume, the higher the variability of results (a bubble plot showing number of hospitals with each annual mortality would have been more informative). As noted in the example cited in the preceding paragraph, this is to be expected. For 1 case, mortality must be 0% or 100%; for 2 cases 0%, 50%, or 100%; and for 3 cases 0%, 33%, 67%, or 100%. As the number of cases increases, the variability diminishes as one gets closer to the unknown "true" underlying mortality of esophagectomy. This is exactly what confidence limits display!

### CONCLUSIONS REACHED AND RECOMMENDATIONS MADE

All this begs the most important issue the authors of these studies address, implicitly or explicitly: Is the volume–mortality relationship an association or causal? Both studies find a statistically significant relationship. If it is causal, then these studies provide some evidence that regionalization may save lives. If it is not causal, regionalization may or may not be of value.\(^2\)
Both sets of authors reject a causal relationship but make different recommendations. Rodgers and colleagues\(^4\) recommend adopting a national benchmark system, which they believe may be a strategy for improving quality of health care and patient outcome. Their study does not provide any evidence for this. In contrast, Meguid and colleagues\(^3\) simply dismiss the statistically significant result and interpret it as “an imperfect surrogate for other variables, which may better define centers of excellence.” A careful look at literature supporting volume–mortality associations in any field of medicine reveals a weak relation. No sharp volume thresholds have been found. So we philosophically side with Meguid and colleagues\(^3\) that using hospital volume for evidence-based hospital selection is misguided.

**EPILOGUE**

These two studies and the literature have demonstrated a weak association between volume and mortality for esophagectomy. However, it is an association with no evidence of cause and effect. Despite this, volume threshold to define Centers of Excellence has become a cause directed at regionalization of care. This strategy is promoted in the guise of improving patient outcome, without evidence that it can do so. It is a strategem.

Lost in striving for high outcome ranking, coveted stars, or Center of Excellence designation are cogent arguments for regionalization of rarely performed complex operations, such as esophagectomy. High volume is essential for training thoracic surgeons, because building skill demands repetition. High volume is necessary for surgeons to keep their skills honed. High volume is essential for efficient cost-effective use of high-tech diagnostic instrumentation. High volume is necessary to generate and maintain an infrastructure of gastrointestinal specialists, oncologists, surgeons, pathologists, radiation therapists, and the like, who can form an effective multidisciplinary team for treating patients with esophageal disease.

Nevertheless, every effort should be made to reduce mortality of esophagectomy toward zero everywhere. Large programs or collaborative groups may discover, by intense statistical analysis or clinical trials, ways to minimize mortality. However, these could just as well come from a low-volume center by a surgeon with keen insight.

**References**

1. Leapfrog Group. [http://www.leapfroggroup.org/media/file/Leapfrog-Evidence-Based_Hospital_Referral_Fact_Sheet.pdf](http://www.leapfroggroup.org/media/file/Leapfrog-Evidence-Based_Hospital_Referral_Fact_Sheet.pdf).


