

Differences in reported esophageal cancer resection outcomes between national clinical and administrative databases

Damien J. LaPar, MD, MSc, George J. Stukenborg, MA, PhD, Christine L. Lau, MD, David R. Jones, MD, and Benjamin D. Kozower, MD, MPH

Background: The Society of Thoracic Surgeons (STS) General Thoracic Surgery Database (GTSD) is the largest clinical thoracic surgical database in the United States. The purpose of the present study was to determine whether the GTSD esophagectomy outcomes are representative of nationwide outcomes by comparing them with other national clinical and administrative databases.

Methods: From 2002 to 2008, esophageal cancer resection outcomes from the GTSD were compared with those from the National Surgery Quality Improvement Program (NSQIP) and Nationwide Inpatient Sample (NIS). The observed differences in patient characteristics and postoperative events were also analyzed.

Results: Annual esophageal resection volumes have increased over time. However, as of 2008, the GTSD and NSQIP only capture a small proportion of resections performed nationally (36% and 11%, respectively). The median patient age and female gender were similar in all 3 databases. Mortality was significantly lower within the GTSD (3.2%) and NSQIP (2.6%) compared with the NIS (6.1%, $P < .001$). The median length of stay was lower in the GTSD (10 days) than in either the NSQIP (12 days) and NIS (12 days, $P < .001$).

Conclusions: The STS GTSD reports outstanding mortality results and hospital length of stay for esophageal cancer resection. However, the surgical outcomes from the STS GTSD are not representative of the national results from programs not participating in the database. These results establish a reference for future esophagectomy comparisons and highlight the importance of increased participation and use of the STS GTSD. (*J Thorac Cardiovasc Surg* 2012;144:1152-9)

Esophageal cancer has accounted for an increasing number of cancer-related deaths in the United States during the past few decades. According to the American Cancer Society, esophageal cancer affects approximately 5 in 100,000 US citizens.¹ Despite improvements in surgical outcomes, recent reports of morbidity and mortality have ranged from 10% to 40% and 0% to 15%, respectively.² As patient outcomes improve, the need for consistent clinical documentation and data gathering are required to help guide future treatment decisions and clinical research.

Significant progress has been achieved in the documentation of patient characteristics and outcomes within both administrative and clinical databases. These databases have facilitated substantial benefits by improving patient care and guiding clinical investigation. However, the content and purpose of these various databases have significant

differences. Currently, the Nationwide Inpatient Sample (NIS) serves as the largest, most representative and publicly available administrative database within the United States.³ Consisting of inpatient discharge records, the NIS reports patient- and hospital-level details and disposition and mortality data. Using the “International Classification of Diseases, Ninth Revision, Clinical Modifications” (ICD-9-CM)-based coding system, the NIS provides an estimate of nationwide trends for patients undergoing medical and surgical admissions within the United States. Among the clinical databases, the American College of Surgeons (ACS) National Surgical Quality Initiative Program (NSQIP) and the Society of Thoracic Surgeons (STS) General Thoracic Surgery Database (GTDB) represent 2 leading surgical databases specifically designed to capture and report general surgery and thoracic surgery outcomes, respectively. The ACS NSQIP was established in 2005 and is a nationally validated, risk-adjusted, outcomes-based program intended to quantify the surgical outcomes and metrics of quality surgical care. Approximately one half of the participating centers in the NSQIP are community hospitals.⁴

The STS GTDB was initiated in 2002 and has demonstrated excellent surgical outcomes for esophageal cancer resection.⁵ The STS GTDB is the largest clinical thoracic surgical database in the United States. However, it is not known whether the outstanding outcomes for esophageal cancer resection from the GTDB are representative of

From the Division of Thoracic and Cardiovascular Surgery, University of Virginia School of Medicine, Charlottesville, Va.

Disclosures: Authors have nothing to disclose with regard to commercial support.

Read at the 92nd Annual Meeting of The American Association for Thoracic Surgery, San Francisco, California, April 28-May 2, 2012.

Received for publication May 1, 2012; revisions received July 24, 2012; accepted for publication Aug 1, 2012; available ahead of print Aug 31, 2012.

Address for reprints: Benjamin D. Kozower, MD, MPH, Division of Thoracic and Cardiovascular Surgery, University of Virginia School of Medicine, PO Box 800679, Charlottesville, VA 22908 (E-mail: bdk8g@virginia.edu).

0022-5223/\$36.00

Copyright © 2012 by The American Association for Thoracic Surgery

<http://dx.doi.org/10.1016/j.jtcvs.2012.08.010>

Abbreviations and Acronyms

ACS	= American College of Surgeons
GTDB	= General Thoracic Surgery Database
ICD-9-CM	= “International Classification of Diseases, Ninth Revision, Clinical Modifications”
NIS	= Nationwide Inpatient Sample
NSQIP	= National Surgical Quality Initiative Program
STS	= Society of Thoracic Surgeons

nationwide results. The purpose of the present study was to determine whether the GTDB esophagectomy outcomes are representative of the nationwide outcomes by directly comparing them with those reported in other clinical and administrative databases. We hypothesized that the GTDB would only capture a few of the procedures performed nationally and that it would report superior outcomes compared with those reported in the NSQIP and NIS databases.

MATERIALS AND METHODS

Data obtained for the present study were extracted from the NIS and STS GTDB for 2002 to 2008. The ACS NSQIP did not begin until 2005, and data were extracted for 2005 to 2008. The present study was exempt from formal review by the University of Virginia institutional review board, because they deemed it was not human subjects research owing to the lack of discrete patient identifiers and because the analyzed data were not exclusively collected for research purposes. The patient records for those undergoing esophageal resection with a diagnosis of esophageal cancer were extracted from each database.

Patient Selection

Within the NIS, appropriate discharge records were identified by querying the first 5 diagnosis and procedure categories using the following ICD-9-CM procedure and diagnosis codes: primary esophageal cancer (ICD-9-CM diagnosis code 150.X) and esophageal resection (ICD-9-CM procedure codes 42.4, 42.40, 42.41, 42.42, and 43.99). The ACS NSQIP and STS GTDB were queried for all patient records for those undergoing primary esophageal cancer resection during the study period. Data were excluded for patients with benign esophageal disease. During data extraction from the NIS data sets, fewer than 1% of patient records with missing data for patient age, gender, and mortality underwent case wise deletion to obtain a complete data set for analysis.

Variable and Outcome Definitions

All variable definitions conformed to those used within the NIS, NSQIP, and GTDB.^{3,4,6} All measured outcomes were established a priori before data analysis. The primary outcomes of interest were the differences in the reported frequency of esophageal cancer resection among the 3 databases and the observed differences in discharge mortality and hospital length of stay. Mortality was defined differently among these 3 databases. The NIS and NSQIP report in-hospital mortality and the GTDB reports 30-day mortality.

Statistical Analysis

All statistical methods were designed to test the null hypothesis that the outcomes after esophageal cancer resection would not be significantly

different. Standard statistical significance was set to an α of 0.05. Descriptive statistics for all variable comparisons were calculated using appropriate univariate hypothesis tests. Categorical variables are expressed as within group percentages and were compared for independent samples using either Pearson's χ^2 or Fisher's exact test. Continuous variables are expressed as either the mean \pm standard deviation or median and interquartile range, depending on the overall variable distribution. Independent-sample, single-factor analysis of variance was used for parametric data comparisons, and the Mann-Whitney U test was used for all nonparametric data comparisons, where appropriate. Calculated test statistics were used to derive all reported 2-tailed P values. All statistical analyses were performed using Predictive Analytics SoftWare (PASW), version 18 (IBM Corp, Somers, NY).

RESULTS

The total number of esophageal cancer resections varied widely among the 3 databases (NIS, 22,758; GTDB, 6740; NSQIP, 1030; Table 1). The median patient age was 63 years in the GTDB, 65 years in the NSQIP, and 64 years in the NIS ($P < .001$). Women underwent a lower frequency of esophageal resection than men in each database. Table 1 also illustrates significant differences in comorbidity among the databases. Of note, the GTDB had a greater prevalence of chronic obstructive pulmonary disease, congestive heart failure, and renal failure than did the NSQIP, the other clinical database included for comparison ($P < .001$).

When examining the annual volume of esophageal cancer resections, the volume of procedures reported increased within all 3 databases (Figure 1). The NIS volumes increased from 2822 operations annually in 2002 to 3728 operations annually in 2008. Esophageal cancer resection frequency consistently increased during the study period within the GTDB and NSQIP but as of 2008 only represented a small proportion (36% and 11%) of the estimated resections performed annually nationwide (Figure 1).

When comparing the primary outcomes, the GTDB reported superior mortality and hospital length of stay compared with the NIS administrative data (Table 2). Discharge mortality in the GTDB and NSQIP was 3.2% and 2.6%, respectively, which was significantly lower than discharge mortality reported within the NIS (6.1%, $P < .001$). In addition, the trends in mortality were lower ($P < .001$) in the GTDB and NSQIP for each operative year throughout the study period (Figure 2). The median hospital length of stay for the GTDB was 2 days less than that in the NSQIP and NIS (GTDB, 10 days vs NSQIP, 12 days vs NSQIP, 12 days; $P < .0001$).

DISCUSSION

We present an important nationwide comparison of the performance of esophageal cancer resection and outcomes as reported within 2 clinical databases, the STS GTDB and the ACS NSQIP, and the most broadly representative administrative database, the NIS. Our results have demonstrated that significant differences exist in reported patient characteristics, esophageal cancer resection frequency,

TABLE 1. Patient demographics and risk factors for esophageal cancer resection as reported in 3 databases

Factor	GTDB (n = 6740)	NSQIP (n = 1030)	NIS (n = 22,758)	P value
Age (y)				.04
Median	63	65	64	
Interquartile range	55-70	56-72	56-72	
Gender				<.0001
Male	82.0%	79.0%	83.0%	
Female	18.0%	21.0%	17.0%	
Year of surgery				<.0001
2002	152	—	2822	
2003	514	—	3058	
2004	579	—	2487	
2005	614	52	3572	
2006	887	209	3074	
2007	1162	367	4017	
2008	1349	402	3728	
Comorbid disease				
Hypertension	48.3%	50.9%	40.6%	<.0001
Heart failure	2.1%	0.2%	4.8%	<.0001
Chronic lung disease	12.7%	6.7%	20.7%	<.0001
Renal failure	1.6%	0.2%	2.2%	<.0001

GTDB, General Thoracic Surgery Database; NSQIP, National Surgical Quality Initiative Program; NIS, Nationwide Inpatient Sample.

and outcomes, as reported by these 3 databases. The number of annual resections reported by the GTDB and NSQIP consistently increased over time. However, the number of esophageal cancer resections reported within both clinical databases only represented a few of the estimated nationwide esophagectomies, as reported by the NIS. Postoperative mortality was significantly better for the clinical databases, and the GTDB reported the lowest hospital

length of stay. These results have significant implications for on the current and future utility of the GTDB compared with the other available clinical and administrative databases as instruments for clinical practice and investigation.

The GTDB is a leading clinical and academic resource within thoracic surgery, providing an important platform from which to perform comparative effectiveness studies and surgical quality assessment. Several GTDB-based studies have recently analyzed and tracked operative outcomes after lung and esophageal resection.^{5,7-11} Among the esophageal cancer resection data, the GTDB has recently been used to identify predictors of major morbidity and mortality and to compare hospital performance variation.⁵ In their analysis of the GTDB, Wright and colleagues⁵ examined more than 2315 esophageal cancer resection cases from 73 different participating centers from 2002 to 2007. Similar to the present findings, their results demonstrated lower hospital mortality (2.7%) than that reported by administrative databases. In their study, the most significant predictors of major morbidity and mortality after esophagectomy were female gender, current cigarette use, and patient comorbidities. Their analyses also demonstrated significant hospital performance variation in risk-adjusted rates of major morbidity and mortality after esophagectomy. Of note, the differences in hospital performance variation were not associated with hospital procedure volume. Wright and colleagues⁵ evaluated this relationship using a participant's average annual procedure volume and, despite a large variation in hospital procedure volume, it was not a significant predictor of mortality or major morbidity. This finding is very similar to recent work reported by Kozower and Stukenborg¹² examining esophageal cancer resection in the NIS. Their study used administrative data

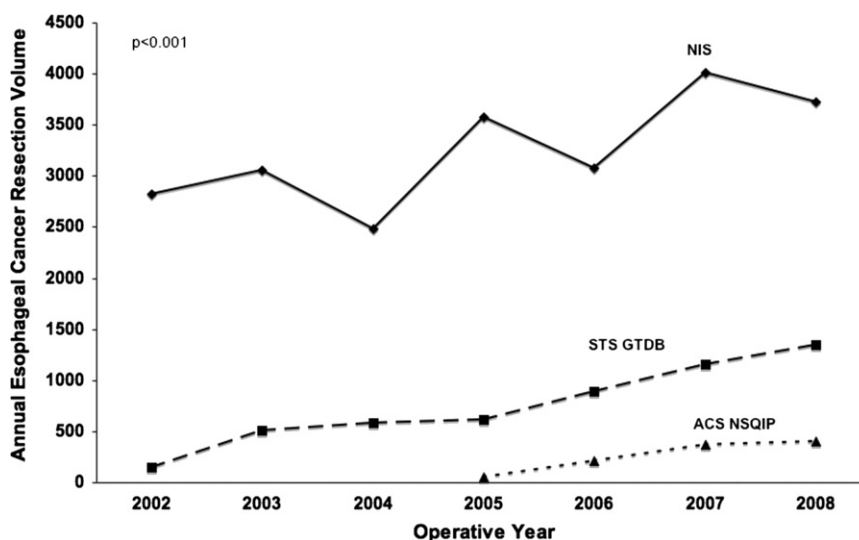


FIGURE 1. Annual esophageal cancer resection volume recorded in General Thoracic Surgery Database (GTDB), National Surgical Quality Initiative Program (NSQIP), and Nationwide Inpatient Sample (NIS) databases ($P < .001$). ACS, American College of Surgeons; STS, Society of Thoracic Surgeons.

TABLE 2. Postoperative morbidity, mortality, and resource use after esophageal cancer resection as reported in 3 databases

Outcome	GTDB (n = 6740)	NSQIP (n = 1030)	NIS (n = 22,758)	P value
Mortality	3.2%	2.6%	6.1%	<.0001
Length of stay (d)				<.0001
Median	10	12	12	
Interquartile range	8-15	9-19	9-19	

GTDB, General Thoracic Surgery Database; NSQIP, National Surgical Quality Initiative Program; NIS, Nationwide Inpatient Sample.

and evaluated the relationship between in-hospital mortality and hospital volume using 3 different techniques for measuring hospital volume. Similar to the clinical study using STS data, Kozower and Stukenborg¹² demonstrated that the hospital esophagectomy volume was not a predictor of mortality and that the most important predictors of mortality were patient age and comorbidities.

A principle finding of the present study was the striking differences in the reported frequencies of esophageal cancer resection among the databases. The NIS was used in the present analysis to provide a valid nationwide estimate of esophageal cancer resection rates from which to compare the GTDB and NSQIP. To ensure a valid analysis of the data contained within each database, comparisons were made for years with available data from both the GTDB and the NIS (2002-2008) and for the years in which the NSQIP had available data (2005-2008). The reported frequency of annual esophageal cancer resection rates within the NIS in the present study was similar to that reported in other published series.^{13,14} A recent series from Kohn and colleagues¹⁴ reported on national trends and outcomes among 57,676 patients undergoing esophagectomy within

the NIS from 1998 to 2006. Connors and associates¹³ reported similar esophageal resection volumes in their comparison of transthoracic and transhiatal esophagectomy among 17,395 patients from 1999 to 2003. With respect to the reported NSQIP volume, Dhungel and colleagues¹⁵ recently reported resection volumes equal to that of the present analysis in their identification of predictors of morbidity and mortality after esophageal resection. Thus, the important finding that, as of 2008, the GTDB and NSQIP only represent 36% and 11%, respectively, of estimated esophageal cancer resections performed within the United States annually provides an important context from which to view the reported results.

The reported mortality and hospital length of stay in the present study are similar to those from other published esophageal cancer resection reports.^{2,5,13-16} In the present series, the hospital discharge mortality rates ranged from 2.6% in the NSQIP and 3.2% in the GTDB to 6.1% in the NIS. Other nationwide series have documented mortality rates for esophageal cancer resection of 6% to 12%,^{13,14} and smaller, single-institution mortality rates have ranged from 0.0% to 15%.^{2,17} Regarding the hospital length of stay, current US nationwide reports for esophagectomy are lacking in recently reported data. However, the results of the present study have demonstrated a median length of stay of 10 to 12 days, comparable to other single-institution reports and meta-analyses.¹⁷⁻¹⁹ More importantly, the principal findings in the present analysis were the significant differences found between the clinical and administrative database outcomes and the low mortality rate and reduced hospital length of stay reported within the GTDB. The significance of the outstanding outcomes and the proportionately small percentage of

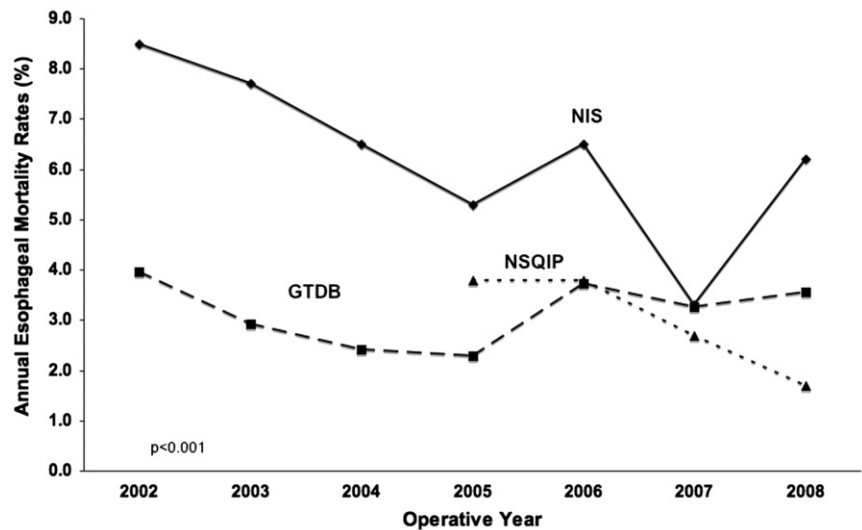


FIGURE 2. Annual mortality after esophageal cancer resection as recorded in General Thoracic Surgery Database (GTDB), National Surgical Quality Initiative Program (NSQIP), and Nationwide Inpatient Sample (NIS) databases ($P < .001$).

cases captured within the GTDB is that these outcomes are not currently generalizable to all patients in the United States. A likely explanation for these differences is that clinical databases include a greater proportion of outcomes from dedicated, experienced thoracic and/or general surgeons, often practicing at tertiary care centers. Moreover, it should be noted that the clinical implications of the slightly longer hospital length of stay reported for patients within the NSQIP and NIS compared with the GTDB likely impart a small effect on overall resource usage, because most costs associated with hospitalizations occur from complications commonly occurring in the first few postoperative days.²⁰

Although the GTDB might not ever represent “generalized” outcomes throughout the United States for nonparticipating, nonthoracic surgeons, we do believe that increased participation in the GTDB is needed to establish its place among available clinical registries and to provide a more accurate measure of outcomes for participating centers. These data demonstrate that future analyses of GTDB esophagectomy outcomes should recognize the context from which these outcomes should be considered. Furthermore, because of the present results, future comparisons of GTDB outcomes with other surgical databases, including the NSQIP, and reported nationwide outcomes are warranted to establish its evolving generalizability, and comparisons of future predictive modeling using these clinical and administrative databases might be feasible.²¹ In addition, recent work examining the generalizability of the GTDB in lung cancer resection showed similar results with the GTDB, representing a small percentage of procedures performed nationally (NIS) but having significantly lower mortality and hospital length of stay.²² These data suggest that the reported superior GTDB outcomes, reflecting the expertise of trained, dedicated thoracic surgeons at tertiary referral hospitals, should not be extrapolated to other general and/or thoracic surgeons and centers not participating in the GTDB.

The reported results have significant clinical implications to support increased participation in the GTDB by US hospitals and thoracic surgeons. Since its introduction, there has been a consistent increase in the number of GTDB participating centers. This increase has been exponential with a dramatic increase in participation within the past 3 to 5 years (Figure 3). This increased participation is directly reflected by the increase in esophageal cancer resections performed during the study period. More importantly, however, the superior results of the GTDB we have reported highlight the critical importance of increased participation in the GTDB to not only allow for improved generalizability of its nationwide outcomes, but also to provide a better benchmark for esophagectomy outcomes by well-trained surgeons and experienced centers.

The present study had several limitations. The retrospective study design is subject to inherent selection bias. A

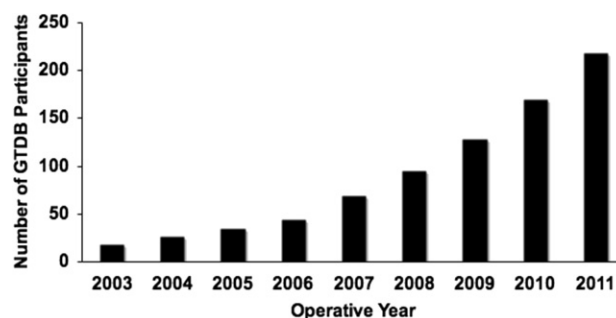


FIGURE 3. Annual number of General Thoracic Surgery Database (GTDB) hospital participants.

second limitation was that different definitions for patient and outcome variables were used in the GTDB, NSQIP, and NIS. Mortality is defined as in-hospital death in the NIS and NSQIP and as 30-day mortality by the GTDB. Also, the 30-day mortality is likely to be lower than in-hospital mortality, and this comparison might underestimate the outstanding results of the GTDB.²³ Additional differences in data content among the 3 databases limited comparisons of specific predictors of surgical outcomes. Detailed patient characteristics remain well-established risk factors for morbidity and mortality after esophageal cancer resection, and many are included in the STS GTDB mortality risk models. Although no precise method is available to determine the specialty practice of surgeons included in these databases, the assumption that a larger percentage of general surgeons are represented in the NSQIP database should be considered. The potential for unrecognized miscoding and data entry errors should be considered in any secondary analysis of data registries. Furthermore, these databases could have some overlap, such that the same patient could be included in more than 1 database. Although external auditing of the GTDB began in 2010, it was not performed for the included study period. However, auditing of the STS adult cardiac database has demonstrated 95% accuracy. Finally, long-term follow-up data are lacking in these databases. Despite these limitations, our results have provided an important contribution to the existing esophagectomy data by providing a benchmark for future GTDB and other database comparisons.

CONCLUSIONS

The present study has demonstrated that significant differences exist between annual procedure volumes and outcomes for esophageal cancer resection reported in clinical and administrative databases. The STS GTDB has reported outstanding mortality results and reduced hospital length of stay for esophageal cancer resection. However, the surgical outcomes from the STS GTDB are not representative of the national results from programs not participating in the database. These results establish a reference for future

esophagectomy comparisons and highlight the importance of increased participation and use of the STS GTDB.

We would like to thank Donna McDonald at the Society of Thoracic Surgeons and Amelia Wallace at the Duke Clinical Research Institute for their assistance in data collection and analysis.

References

1. American Cancer Society. Esophageal cancer. Available from: <http://www.cancer.org/acs/groups/content/@nho/documents/document/esophagealcancerpdf.pdf>. Accessed September 9, 2011.
2. Blencowe NS, Strong S, McNair AG, Brookes ST, Crosby T, Griffin SM, et al. Reporting of short-term clinical outcomes after esophagectomy: a systematic review. *Ann Surg*. 2012;255:658-66.
3. Koch CG, Khandwala F, Nussmeier N, Blackstone EH. Gender and outcomes after coronary artery bypass grafting: a propensity-matched comparison. *J Thorac Cardiovasc Surg*. 2003;126:2032-43.
4. ACS-NSQIP Program Specifics, Surgical Case Inclusion/Exclusion Overview. Available from: http://acsnsqip.org/main/program_case_inclusion_exclusion.asp. Accessed September 9, 2011.
5. Wright CD, Kucharczuk JC, O'Brien SM, Grab JD, Allen MS. Predictors of major morbidity and mortality after esophagectomy for esophageal cancer: a Society of Thoracic Surgeons General Thoracic Surgery Database risk adjustment model. *J Thorac Cardiovasc Surg*. 2009;137:587-96.
6. Vohra HA, Kanwar R, Khan T, Dimitri WR. Early and late outcome after off-pump coronary artery bypass graft surgery with coronary endarterectomy: a single-center 10-year experience. *Ann Thorac Surg*. 2006;81:1691-6.
7. Boffa DJ, Allen MS, Grab JD, Gaisert HA, Harpole DH, Wright CD. Data from the Society of Thoracic Surgeons General Thoracic Surgery database: the surgical management of primary lung tumors. *J Thorac Cardiovasc Surg*. 2008;135:247-54.
8. Kozower BD, Sheng S, O'Brien SM, Liptay MJ, Lau CL, Jones DR, et al. STS database risk models: predictors of mortality and major morbidity for lung cancer resection. *Ann Thorac Surg*. 2010;90:875-83.
9. Onaitis M, D'Amico T, Zhao Y, O'Brien S, Harpole D. Risk factors for atrial fibrillation after lung cancer surgery: analysis of the Society of Thoracic Surgeons general thoracic surgery database. *Ann Thorac Surg*. 2010;90:368-74.
10. Shapiro M, Swanson SJ, Wright CD, Chin C, Sheng S, Wisnivesky J, et al. Predictors of major morbidity and mortality after pneumonectomy utilizing the Society of Thoracic Surgeons General Thoracic Surgery Database. *Ann Thorac Surg*. 2010;90:927-35.
11. Wright CD, Gaisert HA, Grab JD, O'Brien SM, Peterson ED, Allen MS. Predictors of prolonged length of stay after lobectomy for lung cancer: a Society of Thoracic Surgeons General Thoracic Surgery Database risk-adjustment model. *Ann Thorac Surg*. 2008;85:1857-65.
12. Kozower BD, Stukenborg GJ. Hospital esophageal cancer resection volume does not predict patient mortality risk. *Ann Thorac Surg*. 2012;93:1690-6.
13. Connors RC, Reuben BC, Neumayer LA, Bull DA. Comparing outcomes after transthoracic and transhiatal esophagectomy: a 5-year prospective cohort of 17,395 patients. *J Am Coll Surg*. 2007;205:735-40.
14. Kohn GP, Galanko JA, Meyers MO, Feins RH, Farrell TM. National trends in esophageal surgery—are outcomes as good as we believe? *J Gastrointest Surg*. 2009;13:1900-12.
15. Dhungel B, Diggs BS, Hunter JG, Sheppard BC, Vetto JT, Dolan JP. Patient and peri-operative predictors of morbidity and mortality after esophagectomy: American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), 2005-2008. *J Gastrointest Surg*. 2010;14:1492-501.
16. Dikken JL, Dassen AE, Lemmens VE, Putter H, Krijnen P, van der Geest L, et al. Effect of hospital volume on postoperative mortality and survival after oesophageal and gastric cancer surgery in the Netherlands between 1989 and 2009. *Eur J Cancer*. 2012;48:1004-13.
17. Suzuki H, Gotoh M, Sugihara K, Kitagawa Y, Kimura W, Kondo S, et al. Nationwide survey and establishment of a clinical database for gastrointestinal surgery in Japan: targeting integration of a cancer registration system and improving the outcome of cancer treatment. *Cancer Sci*. 2011;102:226-30.
18. Markar SR, Karthikesalingam A, Thrumurthy S, Low DE. Volume–outcome relationship in surgery for esophageal malignancy: systematic review and meta-analysis 2000-2011. *J Gastrointest Surg*. 2012;16:1055-63.
19. Reavis KM, Smith BR, Hinojosa MW, Nguyen NT. Outcomes of esophagectomy at academic centers: an association between volume and outcome. *Am Surg*. 2008;74:939-43.
20. Dimick JB, Pronovost PJ, Cowan JA, Lipsett PA. Complications and costs after high-risk surgery: where should we focus quality improvement initiatives? *J Am Coll Surg*. 2003;196:671-8.
21. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ*. 2007;334:1044-7.
22. LaPar DJ, Bhamidipati CM, Lau CL, Jones DR, Kozower BD. The Society of Thoracic Surgeons General Thoracic Surgery Database: establishing generalizability to national lung cancer resection outcomes. *Ann Thorac Surg*. 2012;94:216-21.
23. Bryant AS, Rudemiller K, Cerfolio RJ. The 30- versus 90-day operative mortality after pulmonary resection. *Ann Thorac Surg*. 2010;89:1717-23.

Discussion

Dr Antoon Lerut (Leuven, Belgium). Congratulations on this excellent presentation and thank you for sending me the manuscript well in advance.

I consider this type of study as being a typical sort of playground for investigators with a master's degree in public health or holding some other master's degree in science, all of which I am not. That is the only disclosure I have to make. Also, I noticed that this is the fifth presentation in this meeting of a report that to some extent relates to the topic of volume and outcomes. That is about 10% of all presentations, which I think comes close to the definition of an epidemic. But, anyway, I will give it a try.

My first observation is that the number of surgeries for esophageal cancer registered, and I presume these are the factual surgeries, are increasing steadily in all databases. The second observation is that the mortality is decreasing slowly, and except for the year 2008 that was in the report, this figure is most prominent in the NIS database, decreasing from 8% in 2002 to 3% in 2007, closing the gap between the NIS and the 2 other databases. I would say that is good news for the patients and also for the surgeons.

I have 3 questions. First, you indicated that comorbidities, such as chronic obstructive pulmonary disease (COPD), which are seen as risk factors for esophageal cancer resection were greater in the GTDB than in the NSQIP database. I wonder whether these databases use the same definitions for the values related to each of the comorbidities. For example, COPD Global Initiative for Chronic Obstructive Lung Disease class 1 and COPD Global Initiative for Chronic Obstructive Lung Disease class 2 both match the definition of COPD but clearly represent differing severity of the disease. That could, of course, be a source a bias. In other words, are you comparing the same severity of disease for each of these comorbidities?

Dr Kozower. Thank you very much for your comments. As you correctly point out, the definitions for various comorbidities, such as COPD, are different for the different databases. We were careful to point this out and tried not to overinterpret the data.

Dr Lerut. The second question relates to the fact that for about 2 decades the relationship between procedure volume and better outcome has been accepted; however, more recently, it has been postulated that volume per se does not translate into better outcomes. Rather, it is the factors associated with it, such as surgeon expertise, perioperative management, and so forth. Therefore, procedure volume should not be used as a measure of quality, as Dr LaPar from your group so strongly advocated last Friday at

the American Surgical Association meeting. I understand that both the GTDB and the NSQIP database include a number of low-volume centers. Because both databases show a very low mortality rate, this seems to confirm that volume on its own is an inadequate parameter of quality measurement. Nevertheless, we keep hearing that esophagectomy needs to be centralized to high-volume centers, and I think all of us would like to be operated for esophageal carcinoma in a high-volume center. So, if it is not volume, what is it that we should bring forward to endorse such a plea in favor of centralization? What do we have to advise to our healthcare policymakers, because they are using that type of data.

Dr Kozower. Your question is obviously a very important one, but you are asking me for kind of the “Holy Grail” of quality. However, we do not fully understand what distinguishes high- and low-performing hospitals. I think expertise is obviously extremely important and quality is a multidimensional concept. So what we have argued so strongly against is trying to oversimplify it and simply use volume as a proxy for quality, particularly when that alone will be used to direct patient referrals. However, factors are present at these high-volume hospitals that are associated with good outcomes. Thus, our group and others are currently working to identify these factors.

Dr Lerut. Well, I would think that is a lot of work ahead, but I think it is important that we try to find these elements, because that is not what the policymakers and insurance companies are using. Also, as I said, I presume that everybody in this room would like to be operated in a high-volume center rather than in a low-volume center. So volume does matter, right?

Dr Kozower. I do not think the data support using volume as a proxy for quality. Even using the STS GTDB, when you consider Cam Wright’s paper on esophagectomy, volume was not a significant predictor of mortality or major morbidity. Importantly, more than one half of the centers are what are classically thought of as low volume, and the results are more generalizable.

Dr Lerut. You mentioned that the GTDB and the NSQIP database represent only a small proportion, 36% and 11%, respectively, of the estimated number of resections performed nationwide. Given the comparable results, is it not time to link those 2 databases to generate a larger effort and therefore a stronger sample? This would allow us to better convince the outside world that patients should be treated by board-certified surgeons, whether thoracic surgeons or specialist upper gastrointestinal surgeons. I come from a part of the world where I am more frequently confronted with the fact that upper gastrointestinal surgeons are involved in performing this type of surgery with excellent results. I would like to hear your comments on how you see this in this country.

Dr Kozower. While I am not in charge of setting US policy, I strongly believe that it is the surgical outcomes that should be used to judge the surgeon and not any arbitrary predictor. Therefore, it is going to be very important for us to determine how to measure and differentiate surgical and hospital performance. The STS database has demonstrated significantly better outcomes, and so it is not hard to convince this crowd that esophageal cancer should be performed by those of us sitting in the room. How do you do that for the whole country, and are we actually available to the whole country? These are very complicated issues. Canada is starting to figure this out, and they have been presenting their series with regionalization of care. Some of the results seem quite

good and some of them actually have not shown as large a decrease in mortality as one might think.

Dr Lerut. Would you make an effort to combine the 2 databases and try to come up with a greater percentage and therefore be able to make stronger statements on just the idea to centralize patients?

Dr Kozower. I do not think I would want to combine the 2 databases, because they have very different purposes. The NSQIP was established for general and vascular surgery and uses a sampling method to capture thoracic procedures, such as esophagectomy, performed at a given hospital. Certain hospitals use the NSQIP to capture every case they perform, but, overall, that is not the sampling method.

Dr Lerut. I thank the Association for the privilege to discuss this paper.

Dr David J. Sugarbaker (Boston, Mass). I have 1 question and comment regarding the STS database and NSQIP or the administrative, and that is that I am not sure people are aware that the fundamental difference between the NSQIP and STS database, and one of the disadvantages of the STS database, could be that it uses elective data submission. So, the data submission for STS is retrospective, occurring after the procedure, usually after discharge, and selection bias in any type of protocol in which that is the case is certainly recognized as a problem. The NSQIP is a real-time snapshot of what is happening, and the administrative database is, again, a nonelective submission of data, usually electronically. The STS database is helpful, but it is not an intent-to-treat, prospective registration of patients that in most assessments would give one a clearly accurate assessment of what is occurring. Do you want to comment on that?

Dr Kozower. You are absolutely correct. The only point I would add is that the latest version of the database differentiates between major and minor procedures and requires that all major procedures be reported. External audits are being used to verify the accuracy of the data submitted and to verify that major cases, perhaps with poor outcomes, are not being omitted. The results these audits will obviously be extremely important.

Dr Sugarbaker. I would just say, if you know, particularly in the cardiac world, what pressure institutions are under for outcomes, you could be concerned that in particular centers tremendous pressure exists to produce excellent outcomes, not that anybody is going to alter that, but there is tremendous pressure.

Dr Alexander Krupnick (St. Louis, Mo). Going along with Dr Sugarbaker’s question, is 1 of these databases audited by insurance companies? In other words, does this reporting affect reimbursement? Could some of the comorbidities be overexaggerated in 1 database and reported more accurately in others? Is there any financial incentive for differential reporting between the databases and could that in any way contribute toward the discrepancy?

Dr Kozower. Excellent question. The NIS actually uses the same discharge abstract data that is used to generate charges. So, if there is intentional gamesmanship, it is truly fraud. However, we all know that issues exist with the accuracy of administrative data. The more we study our own administrative data, from the University Healthsystem Consortium, the more issues we find. For the STS database, the current audits have looked quite good. Fortunately, no current advantage exists for “gaming” the system

currently. As Dr Sugarbaker appropriately points out, as the pressure increases and the insurance companies “tighten the screws,” there will be some increased pressure to perform.

Dr Sugarbaker. Particularly because the STS database is now being made available to certain insurers to assess quality for particular institutions regionally.

Dr Thomas K. Waddell (*Toronto, Ontario, Canada*). I had the same concerns, but when I first read this abstract, Benji, and I know you are such a methodological purist, I thought the purpose of this study was to actually try to study the various qualities of different types of databases, not to actually talk about that this proves that this type of surgeon is better to do this operation in this type of hospital.

So, returning to a good question; that is, can we look at the quality of data in these 3 databases, are there actually any overlapping patient data that are in all 3 data sets? Is it possible to perform patient-level linkage? Is it possible to perform hospital-level linkage? That would get to David’s question about what is the actual missing patient who often had, perhaps, died. Also, even to go a little bit more broadly, could you actually construct some sort of quality assessment of the data entered by linking to the Centers for Medicare and Medicaid Services data to study items such as the reoperation

rate as a method to judge the quality of not just the STS database but, in fact, all of these data sets?

Dr Kozower. That is a great question. When I started this project, my initial objective was to compare these different databases at a national level. However, we cannot link them at a patient level to compare the accuracy of data and reported outcomes such as mortality, morbidity, and readmission. Although certain hospital level characteristics can be linked between different databases, patients are deidentified.

I am sorry, I did not fully catch your question on the Centers for Medicare and Medicaid Services.

Dr Waddell. I was wondering if you could study reoperation as an outcome metric of quality esophagectomy, or whatever, and then use the Centers for Medicare and Medicaid Services data to validate that; how many patients actually were there bills or charges for associated with that reoperation. That could be one method to validate. If you find that everybody is billing for 3 times more reoperations than they claim to be performing in the STS database, that might tell you something.

Dr Kozower. It is an excellent question and we could pool data from a few hospitals to consider this, but we cannot do it for the national databases.