Longitudinal Follow-up of Lung Cancer Resection From the Society of Thoracic Surgeons General Thoracic Surgery Database in Patients 65 Years and Older

Felix G. Fernandez, MD, MS, Anthony P. Furnary, MD, Andrzej S. Kosinski, PhD, Mark W. Onaitis, MD, Sunghee Kim, PhD, Daniel Boffa, MD, Patricia Cowper, PhD, Jeffrey P. Jacobs, MD, Cameron D. Wright, MD, and Joe B. Putnam, Jr, MD

Emory University, Atlanta, Georgia; Starr-Wood Cardiac Group, Portland, Oregon; Duke University, Durham, North Carolina; Yale University, New Haven, Connecticut; Johns Hopkins All Children's Heart Institute, St. Petersburg, Florida; Massachusetts General Hospital, Boston, Massachusetts; and Baptist MD Anderson Cancer Center, Jacksonville, Florida

Background. **The Society of Thoracic Surgeons (STS) General Thoracic Surgery Database (GTSD) does not capture long-term survival after lung cancer surgery. Our objective was to provide longitudinal follow-up to the STS GTSD through linkage to Centers for Medicare and Medicaid Services (CMS) data for patients 65 years of age or older.**

Methods. **Lung cancer operations reported in the STS GTSD from 2002 through 2012 were linked to CMS data for patients 65 years of age or older using variables common to both databases with a deterministic matching algorithm. Mortality data were abstracted for each linked patient from the CMS data. The Kaplan-Meier method was used to estimate long-term survival for lung cancer surgery patients based on tumor stage.**

Results. **From 2002 through 2012, 60,089 lung cancer resections were identified in the GTSD, and 37,009 (61.7%) were in patients 65 years or older. Of these, 26,055 of 37,099 lung cancer resections (70%) in patients 65 years or older were successfully linked to CMS data. Failure to link was most commonly related to having a health maintenance organization or commercial insurance as the primary payer: 40.5% (5,290 of 13,065) of such patients were not linked from 2009 to 2012 (years payer data available). Median survival after lung cancer resection was 6.7 years for pathologic stage I, 3.5 years for stage II, 2.4 years for stage III, and 2.2 years for stage IV.**

Conclusions. **The CMS data complement the STS GTSD data by enabling examination of long-term survival and resource utilization in patients 65 years or older. Linked data from the STS GTSD and the CMS will allow for longitudinal analyses of comparative effectiveness among different surgical approaches for the treatment of lung cancer.**

(Ann Thorac Surg 2016;101:2067–76)

The Society of Thoracic Surgeons (STS) General Thoracic Surgery Database (GTSD) was established in 2002 as a component of the STS National Database to support quality improvement efforts in general thoracic surgery [1]. The STS GTSD captures detailed clinical data and outcomes related to the immediate postoperative period. Risk-adjusted results are provided to database participants twice yearly. The STS has also used the GTSD to develop perioperative risk models and perform analyses of comparative effectiveness for various thoracic operations with respect to short-term outcomes. These studies have improved the safety of lung cancer surgery as well as that of other thoracic surgical procedures.

Despite many advantages, the STS GTSD has been limited by the absence of information regarding outcomes beyond the immediate postoperative period of 30 days or hospital discharge. As procedural morbidity and

Accepted for publication March 7, 2016.

Presented at the Fifty-second Annual Meeting of the Society of Thoracic Surgeons, Phoenix, AZ, Jan 23–27, 2016.

Address correspondence to Dr Fernandez, The Emory Clinic, 1365 Clifton Rd, NE, Ste A2214, Atlanta, GA 30322; email: felix.fernandez@emoryhealthcare.org.

The Appendix can be viewed in the online version of this article [http://dx.doi.org/10.1016/j.athoracsur.2016.03.034] on http://www.annalsthoracicsurgery.org.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2016.03.034

mortality decreases, particularly as related to lung cancer surgery, the confirmation of long-term vital status and resource utilization takes on increasing importance. Because many important events occur after the index hospitalization, limited long-term follow-up is a critical barrier to the optimal utilization of registry data. This limitation has been overcome in the STS Adult Cardiac Surgery Database by linkage to Centers for Medicare and Medicaid Services (CMS) data [2].

For any cancer operation, 5-year survival is a significant measure of clinical success. Lung cancer is the leading cause of cancer mortality in men and women in the United States, with an estimated 221,200 new cases and 158,040 deaths expected in 2015 [3]. The absence of longitudinal follow-up limits the ability of the STS GTSD to assess the quality and effectiveness of lung cancer surgery, affecting a significant proportion of these patients. The objective of this study was to provide the first longitudinal follow-up for STS GTSD lung cancer operations through linkage to CMS data. Successful linkage of the STS GTSD to CMS data will allow for analyses of comparative effectiveness on lung cancer surgery, examining long-term survival and economic outcomes.

Patients and Methods

Institutional Review Board Approval

This study was approved by the Duke University Health System Institutional Review Board. The data used in this analysis represent a limited data set with no direct patient identifiers that was originally collected for nonresearch purposes. Therefore, the analysis of these data was declared by the Duke University Health System Institutional Review Board to be research not involving human subjects [4].

The Society of Thoracic Surgeons General Thoracic Surgery Database

The STS GTSD was queried for all patients treated with surgical resection for primary lung cancer from January 1, 2002, through December 31, 2012. This study encompassed several versions of the STS GTSD data collection instruments (v2.07, v2.081, and v2.2), with various changes incorporated to reflect evolutions in thoracic operations. The STS GTSD has been externally audited since 2010 [5]. Audits have demonstrated high agreement rates with hospital records and validated the accuracy and completeness of the data.

Medicare Database of Centers for Medicare and Medicaid Services

Medicare is health insurance provided by the federal government of the United States for the following groups of patients: (1) people 65 years or older, (2) people younger than 65 with certain disabilities, and (3) people of any age with end-stage renal disease. The CMS administrative claims data source for this study is the 100% Medicare inpatient claims file, which contains information on hospitalizations of patients enrolled in fee-for-service Medicare. It includes dates of service and diagnostic codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). The database contains anonymous patient identifiers that enable follow-up of beneficiaries over time, but does not enable identification of any beneficiary through their Medicare health insurance number. The 100% Medicare denominator file, which links to the inpatient file and contains information on beneficiary eligibility, demographic characteristics, and date of death, was also used.

Linkage of The Society of Thoracic Surgeons Records to Centers for Medicare and Medicaid Services Data

The STS GTSD was linked to CMS claims files using combinations of nonunique indirect identifiers through a deterministic matching algorithm [2, 4]. Records in the two databases were considered to contain information on the same patient if they matched on a set of indirect identifiers including hospital, date of birth, sex, admission date, and discharge date. The CMS 100% inpatient claims file was first queried for ICD-9-CM procedure codes for pulmonary resections 32.0 to 32.9. Variables common to both the STS GTSD and the CMS database were then used to link operations reported in the STS GTSD to CMS inpatient claims data for all hospitalizations during which a patient underwent lung cancer resection. Because of the inclusion criteria and the nature of the databases, this study was restricted to patients who underwent surgery from 2002 through 2012, inclusive, who were 65 years of age or older at the time of surgery. Once the individual patients were linked, longitudinal records were created containing follow-up information, including subsequent death. From this follow-up information, it will also be possible to identify and capture readmissions and to obtain data including primary and secondary diagnoses, repeat surgical procedures, mortality, major morbidity, and charges.

Lung Cancer Staging

For each STS lung cancer resection, clinical and pathologic stages were assigned using the American Joint Committee on Cancer (AJCC) 7th edition staging system. Stage assignment was based on available T, N, and M descriptors recorded in the GTSD. The granularity of the T descriptors in the GTSD has changed with time. In database v2.07 (years 2002–2008), T stage was recorded as follows: T1, T2, T3, or T4. Subsequently, in database v2.081 and v2.2, the T stage has more specificity: T1a, T1b, T2a, T2b, T3, and T4. This affected the analysis in that a T2a N0 tumor would be classified stage IB and a T2b N0 tumor would be classified stage IIA. Because the number of T2a N0 tumors in the GTSD (v.2.081 and v2.2) exceeds the number or T2b N0 tumors by a ratio of greater than 4:1, T2 N0 tumors from database version v2.07 were coded as stage I rather than stage II. Sensitivity analyses with T2 N0 tumors from v2.07 coded as stage II were also performed.

Statistical Analysis

Descriptive statistics were generated for three patient cohorts: STS lung cancer resections in patients 65 years or older linked to CMS data, STS lung cancer resections in

patients 65 years or older *not* linked to CMS data, and STS lung cancer resections in patients younger than 65 years. The Kaplan-Meier method was then used to estimate survival after lung cancer surgery in the STS-CMS linked cohort for strata of patients including both the clinical and pathologic tumor stage groupings. Differences between strata were examined with the log rank test. All statistical tests were two-sided, and probability values of 0.05 or less were considered statistically significant. Subgroup analyses were performed according to race, sex, and age. Expected survival curves for selected age groups were created for comparison with survival after resection of stage I lung cancer using the 2002 through 2011 US sex-specific life tables and the Ederer method for computing cohort survival [6, 7]. Analyses were performed using SAS 9.4 (SAS Institute Inc, Cary, NC) and R 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria) statistical packages.

Results

Serial exclusions in creation of the STS-CMS linked lung cancer resection cohort are detailed in the Appendix (eTable 1). From 2002 through 2012, the STS GTSD contains information regarding 112,606 pulmonary operations, of which 60,089 were lung cancer resections; 37,009 (61.7%) were patients 65 years or older. Subsequently, 26,055 of 37,099 lung cancer resections (70%) in patients 65 years or older were successfully linked to CMS data. Failure to link an STS lung cancer resection record to CMS data was most commonly related to having a health maintenance organization or commercial insurance as the primary payer. Of patients with health maintenance organization or commercial insurance, 40.5% (5,290 of 13,065) were not linked from 2009 through 2012 (database v2.081 and 2.2). Payer data were not recorded in database v2.07.

Descriptive statistics of three patient cohorts are shown in Table 1: STS-CMS linked lung cancer resections in patients 65 years or older, STS lung cancer resections in patients 65 years or older not linked to CMS, and STS lung cancer resections in patients younger than 65 years. In patients 65 years or older, cohorts linked and not linked to CMS data were clinically similar with respect to the distributions of age, sex, year of surgery, performance of a forced expiratory volume in 1 second (FEV_1) test, percent predicted FEV_1, use of induction therapy, clinical and pathologic stage, and primary procedure performed. However, because of the large sample size, differences among groups were statistically significant for all variables except FEV_1 test done and induction therapy. There were relatively more African Americans in the nonlinked group. The operative mortality, defined as a death in hospital or within 30 days of the index procedure, was higher in the nonlinked cohort (2.7% versus 2.2%). The cohort of lung cancer resections in patients younger than 65 years was more likely to receive induction therapy, more likely to undergo a pneumonectomy, and had a lower operative mortality compared with patients 65 years or older.

In the STS-CMS linked group, the initial STS-reported operative mortality of 2.2% increased to 2.6% (667 of 26,055) when the STS data were supplemented with mortality data from CMS records. In the STS-CMS linked cohort, 90-day mortality was 4.7%. Mortality at 90 days after lung cancer surgery stratified by extent of pulmonary resection is shown in Table 2 for the STS-CMS linked group. Importantly, the 90-day mortality was at least twice the STS-reported mortality for sublobar and lobar resection and increased to 10% and 15.8% for bilobectomy and pneumonectomy, respectively. However, length of hospital stay in the STS-CMS linked cohort was similar when measured using the STS GTSD or CMS data (6.48 days; 95% confidence interval, 6.4 to 6.57 versus 6.44 days; 95% confidence interval, 6.37 to 6.52, respectively).

Among all linked patients 65 years or older (n = 26,055), median follow-up was 1.93 years (interquartile range, 0.84 to 3.65 years); this represents the time interval from surgery to time of death or administrative censorship. In this group of patients, the time interval from surgery date to the end of the study (December 31, 2012) was a median of 3.06 years (interquartile range, 1.43 to 5.15 years). This represents the time that was potentially available for follow-up. Overall survival based on clinical and pathologic stage for the entire cohort of STS lung cancer resections linked to CMS data (years 2002–2012, database v2.07, v2.081, and v2.2) is shown in Figure 1. Interestingly, there are no significant survival differences between clinical stages II, III, and IV, highlighting the inability of current staging modalities to identify metastases precisely. Median survival after lung cancer resection for pathologic stages I, II, III, and IV was 6.7, 3.5, 2.4, and 2.2 years, respectively.

Survival based on era of operation comparing survival curves for surgical years 2002 through 2008 (database v2.07) and surgical years 2009 through 2012 (database v2.081 and v2.2) based on both clinical and pathologic stage is shown in the Appendix (eFig 1). There were small improvements in survival overtime when eras of surgery were compared. Sensitivity analyses, with T2 N0 tumors classified as stage II for database v2.07, are also shown in the Appendix (eFig 2). Small increases in survival for stages I and II are observed in this sensitivity analysis.

When additional exploratory subgroup analyses were performed for resected stage I lung cancer, unadjusted survival curves demonstrated superior survival for women (Fig 2). No survival difference was noted for resected pathologic stage I lung cancer when examined according to race (Fig 3). Finally, analysis of stage I survival according to age groupings (65–69 years, 70–74 years, 75–79 years, and 80 years or older) demonstrates an expected reduced survival with increasing age at diagnosis (Fig 4). All age groupings have significantly worse survival than expected age-matched control subjects in the general population.

Comment

Lung cancer operations in patients 65 years or older in the STS GTSD were successfully linked to CMS data to overcome the limitation of the STS GTSD data to

Table 1. Descriptive Statistics of The Society of Thoracic Surgeons General Thoracic Surgery Database Lung Cancer Resection Patients 2002–2012[a]

Variable	Overall (n = 60,089)	Age ≥65 y Linked to CMS (n = 26,055)	Age ≥65 y Not Linked to CMS (n = 11,044)	Age <65 y (n = 22,990)
Age (y), median (IQR)	68.0 (60.0, 74.0)	73.0 (69.0, 78.0)	72.0 (68.0, 77.0)	58.0 (52.0, 61.0)
Sex				
Male	28,451 (47.4%)	12,635 (48.5%)	5,614 (50.9%)	10,202 (44.4%)
Female	31,591 (52.6%)	13,407 (51.5%)	5,418 (49.1%)	12,766 (55.6%)
Race				
Caucasian	52,069 (91.0%)	23,344 (93.7%)	9,483 (90.5%)	19,242 (88.2%)
African American	4,895 (8.6%)	1,494 (6.0%)	937 (8.9%)	2,464 (11.3%)
Other	225 (0.4%)	68 (0.3%)	55 (0.5%)	102 (0.5%)
Year of surgery				
2002	537 (0.9%)	240 (0.9%)	90 (0.8%)	207 (0.9%)
2003	1,134 (1.9%)	392 (1.5%)	257 (2.3%)	485 (2.1%)
2004	1,899 (3.2%)	880 (3.4%)	252 (2.3%)	767 (3.3%)
2005	2,491 (4.1%)	1,179 (4.5%)	343 (3.1%)	969 (4.2%)
2006	3,886 (6.5%)	1,771 (6.8%)	549 (5.0%)	1,566 (6.8%)
2007	5,202 (8.7%)	2,367 (9.1%)	832 (7.5%)	2,003 (8.7%)
2008	6,683 (11.1%)	2,919 (11.2%)	1,205 (10.9%)	2,559 (11.1%)
2009	8,121 (13.5%)	3,486 (13.4%)	1,579 (14.3%)	3,056 (13.3%)
2010	9,049 (15.1%)	3,827 (14.7%)	1,723 (15.6%)	3,499 (15.2%)
2011	10,023 (16.7%)	4,387 (16.8%)	1,875 (17.0%)	3,761 (16.4%)
2012	11,064 (18.4%)	4,607 (17.7%)	2,339 (21.2%)	4,118 (17.9%)
FEV_1 test performed	53,651 (89.3%)	23,491 (90.2%)	9,896 (89.6%)	20,264 (88.1%)
FEV_1 % predicted, median (IQR)	79.0 (65.0, 93.0)	79.0 (64.0, 94.0)	78.0 (63.0, 93.0)	80.0 (66.0, 93.0)
Induction therapy				
Chemotherapy only	2,396 (4.0%)	885 (3.4%)	379 (3.4%)	1,132 (4.9%)
Radiation therapy only	533 (0.9%)	248 (1.0%)	93 (0.8%)	192 (0.8%)
Chemotherapy and radiation	3,262 (5.4%)	933 (3.6%)	398 (3.6%)	1,931 (8.4%)
Pathologic stage				
Stage I	38,194 (65.5 %)	17,407 (68.8%)	7,147 (66.4%)	13,640 (61.4%)
Stage II	10,810 (18.6%)	4,412 (17.4%)	2,009 (18.7%)	4,389 (19.8%)
Stage III	7,613 (13.1%)	2,919 (11.5%)	1,301 (12.1%)	3,393 (15.3%)
Stage IV	1,651 (2.8%)	550 (2.2%)	306 (2.8%)	795 (3.6%)
Clinical stage				
Stage I	38,873 (72.6%)	17,687 (76.0%)	7,256 (73.8%)	13,930 (68.2%)
Stage II	7,758 (14.5%)	3,111 (13.4%)	1,412 (14.4%)	3,235 (15.8%)
Stage III	5,562 (10.4%)	2,020 (8.7%)	940 (9.6%)	2,602 (12.7%)
Stage IV	1,360 (2.5%)	460 (2.0%)	227 (2.3%)	673 (3.3%)
Primary procedure				
Wedge resection	9,913 (16.5%)	4,604 (17.7%)	2,179 (19.7%)	3,130 (13.6%)
Segmentectomy	2,991 (5.0%)	1,833 (7.0%)	464 (4.2%)	694 (3.0%)
Lobectomy	40,523 (67.4%)	17,518 (67.2%)	7,427 (67.2%)	15,578 (67.8%)
Sleeve lobectomy	893 (1.5%)	243 (0.9%)	113 (1.0%)	537 (2.3%)
Bilobectomy	1,998 (3.3%)	710 (2.7%)	350 (3.2%)	938 (4.1%)
Pneumonectomy	2,725 (4.5%)	799 (3.1%)	340 (3.1%)	1,586 (6.9%)
Miscellaneous	1,046 (1.7%)	348 (1.3%)	171 (1.5%)	527 (2.3%)
Operative mortality	1,166 (1.9%)	582 (2.2%)	296 (2.7%)	288 (1.3%)

[a] Data are No. (%) unless otherwise indicated. $p < 0.05$ for all variables.

CMS = Centers for Medicare and Medicaid Services; FEV_1 = forced expiratory volume in 1 second; IQR = interquartile range.

in-hospital or 30-day outcomes. With the addition of longitudinal follow-up, survival outcomes that are more relevant to cancer may be examined in patients undergoing lung cancer surgery. Further, access to Medicare claims data will allow cost and resource use outcomes after lung cancer surgery to be evaluated.

The CMS administrative data was selected as the data source for external linkage with the STS GTSD for

Table 2. Operative and 90-Day Mortality Stratified by Primary Procedure[a]

Variable	Wedge Resection (n = 4,604)	Segmentectomy (n = 1,833)	Lobectomy (n = 17,518)	Sleeve Lobectomy (n = 243)	Bilobectomy (n = 710)	Pneumonectomy (n = 799)
Operative mortality: GTSD data only	52 (1.1%)	30 (1.6%)	367 (2.1%)	6 (2.5%)	44 (6.2%)	63 (7.9%)
Operative mortality: composite of GTSD and CMS data	71 (1.5%)	33 (1.8%)	416 (2.4%)	6 (2.5%)	48 (6.8%)	72 (9.0%)
Mortality within 90 days of surgery: composite of GTSD and CMS data	162 (3.5%)	65 (3.5%)	751 (4.3%)	16 (6.6%)	71 (10.0%)	126 (15.8%)

[a] Data are No. (%).

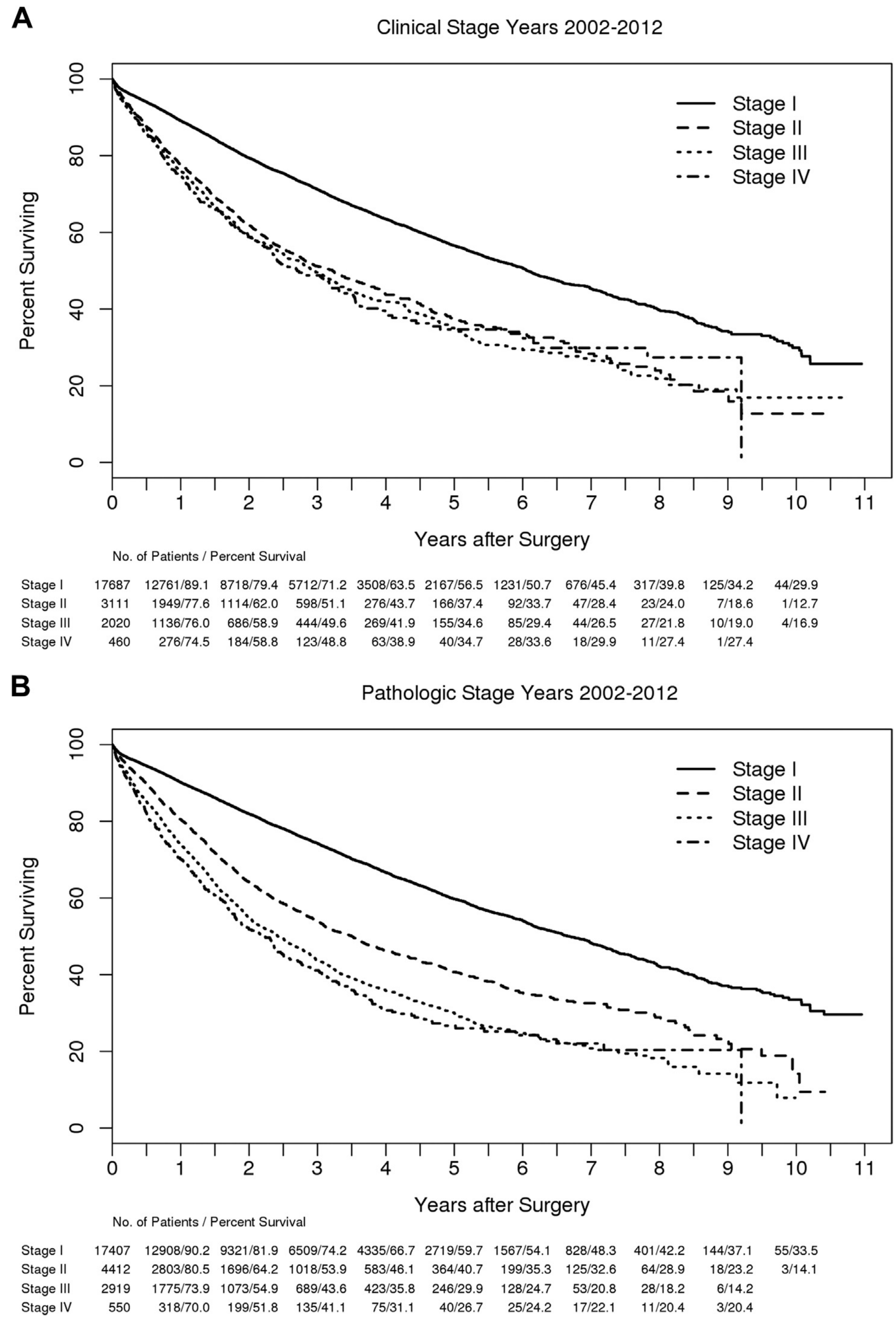
CMS = Centers for Medicare and Medicaid Services; GTSD = General Thoracic Surgery Database.

longitudinal follow-up. This follows the paradigm of the STS Adult Cardiac Surgery Database, which has previously been linked to CMS data [2, 8, 9]. Unfortunately, CMS data provide follow-up only for Medicare beneficiaries, the vast majority of whom are 65 years or older. Linkage to Medicare beneficiaries younger than 65 years was not performed as these patients are small in number and have renal failure or other disabilities. The Social Security Death Master File could provide survival data on the entire cohort of patients in the STS GTSD, making the data generalizable outside the elderly. Unfortunately, in 2011 restrictions to access of the Social Security Death Master File were imposed by the Social Security Administration. Should access to the Social Security Death Master File again become available, this strategy could be implemented. However, the median age of lung cancer surgery patients in the STS GTSD is 68 years, and 61.8% are 65 years or older. Therefore, the STS GTSD-CMS linked dataset is generalizable to a large number of lung cancer surgery patients.

Linkage to CMS data was successful in 70% of lung cancer surgery patients in the STS GTSD 65 years or older using a previously established deterministic matching algorithm [2]. Consequently, longitudinal follow-up was not available for 30% of patients in this cohort. To determine reasons for failure of linkage to CMS data, the primary payer for the lung cancer surgery hospitalization was examined. Forty percent of patients with private insurance or a health maintenance organization as a primary payer during the years 2009 through 2012 could not have their records successfully linked to CMS data. Information on the primary payer is not available in the STS GTSD before 2009 (database v2.07). If claims for the hospital stay related to the lung cancer surgery were not submitted to CMS, linkage would not be feasible. Other exploratory analysis discovered missing data elements in variables used for linkage from certain centers reporting to the STS GTSD. A case-by-case analysis of unlinked records is not feasible owing to the overall volume of records. However, given the small differences noted in Table 1, we believe that our matched cohort is representative of all lung cancer resections in STS-reported patients 65 years or older.

Striking differences are not noted in the demographics of linked and nonlinked patients, although there were a greater percentage of African Americans in the nonlinked cohort (8.9% versus 6.0%). Operative mortality was greater in nonlinked patients as well when using only STS GTSD data (2.7% versus 2.2%). Why operative mortality is higher in patients not linked to CMS data is not known, but merits further examination. It is also interesting to note that in the linked cohort, when STS GTSD data on operative mortality are supplemented by vital status data in CMS files, the operative mortality rate increases from 2.2% to 2.6%. This observation is most likely attributable to more complete capture of postoperative events with the combined dataset; however, selective underreporting must also be considered. Operation-related deaths continue to be observed past the 30-day or hospital discharge period, as 90-day mortality was 4.7% in the linked cohort. Noticeable increases in mortality were observed at 90 days for all lung cancer resection procedures. This important finding has been noted by others [10]. Patients should be provided with these data, and further study of methods to reduce these high numbers, particularly after pneumonectomy, is warranted.

Five-year survival after surgical resection of lung cancer stratified by pathologic stage in the STS-CMS dataset is comparable to that demonstrated by Goldstraw and colleagues [11] in the International Association for the Study of Lung Cancer Lung Cancer Staging Project. Five-year survival based on pathologic stage in the International Association for the Study of Lung Cancer project was as follows: IA, 73%; IB, 58%; IIA, 46%; IIB, 36%; IIIA, 24%; IIIB, 9%; and IV, 13%. In the present study, using STS-CMS data, pathologic stage-specific 5-year survival was observed to be as follows: I, 59.7%; II, 40.7%; III, 29.9%; and IV, 26.7. The greater than expected survival seen for stages III and IV in the STS-CMS data may result from the highly selected nature of these patients with advanced lung cancer treated with surgical therapy. For example, surgery for stage IV lung cancers is undertaken only in exceptional circumstances in the setting of solitary metastases, typically cerebral or adrenal. Additional unadjusted exploratory survival analyses demonstrate that survival in the STS-CMS data varies according to sex and age. These findings are expected and consistent with many data that have previously been reported [12]. Demonstration of these survival curves validates the use of the STS-CMS data for long-term survival analyses in lung cancer surgery.



No. of Patients / Percent Survival

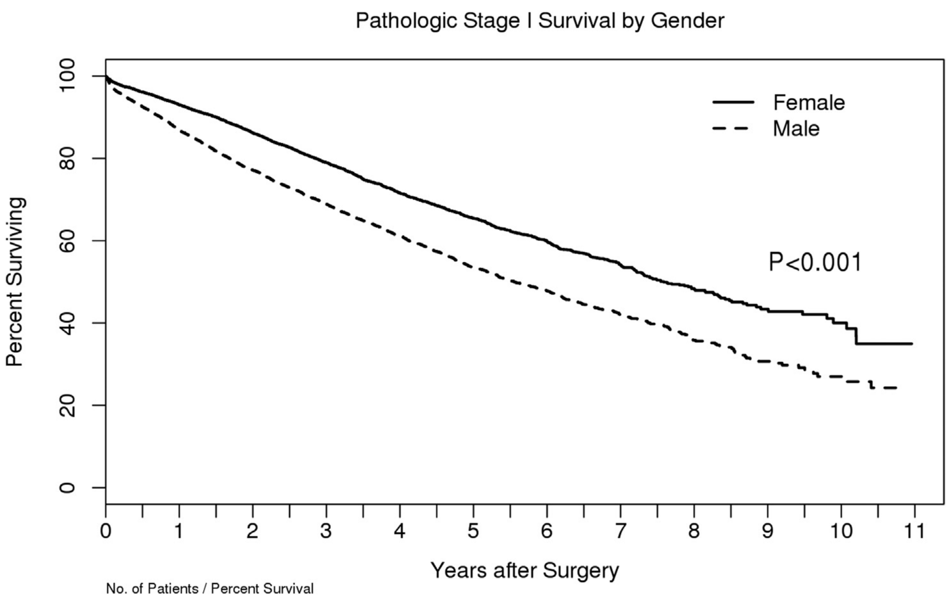
Stage I	17687	12761/89.1	8718/79.4	5712/71.2	3508/63.5	2167/56.5	1231/50.7	676/45.4	317/39.8	125/34.2	44/29.9
Stage II	3111	1949/77.6	1114/62.0	598/51.1	276/43.7	166/37.4	92/33.7	47/28.4	23/24.0	7/18.6	1/12.7
Stage III	2020	1136/76.0	686/58.9	444/49.6	269/41.9	155/34.6	85/29.4	44/26.5	27/21.8	10/19.0	4/16.9
Stage IV	460	276/74.5	184/58.8	123/48.8	63/38.9	40/34.7	28/33.6	18/29.9	11/27.4	1/27.4	

No. of Patients / Percent Survival

Stage I	17407	12908/90.2	9321/81.9	6509/74.2	4335/66.7	2719/59.7	1567/54.1	828/48.3	401/42.2	144/37.1	55/33.5
Stage II	4412	2803/80.5	1696/64.2	1018/53.9	583/46.1	364/40.7	199/35.3	125/32.6	64/28.9	18/23.2	3/14.1
Stage III	2919	1775/73.9	1073/54.9	689/43.6	423/35.8	246/29.9	128/24.7	53/20.8	28/18.2	6/14.2	
Stage IV	550	318/70.0	199/51.8	135/41.1	75/31.1	40/26.7	25/24.2	17/22.1	11/20.4	3/20.4	

Fig 1. Overall survival for surgically managed primary lung cancer in patients 65 years or older with The Society of Thoracic Surgeons General Thoracic Surgery Database records linked to Centers for Medicare and Medicaid Services data based on (A) clinical stage and (B) pathologic stage. Stage I is indicated by the solid line; stage II is indicated by the dashed line; stage III is indicated by the dotted line; and stage IV is indicated by the dash-dot line.

The STS-CMS linked dataset represent a unique resource for outcomes research in lung cancer surgery. Traditional models of survival in lung cancer patients are based on registry data containing tumor stage but limited clinical detail. Surveillance, Epidemiology and End Results data; Surveillance, Epidemiology and End Results-Medicare linked data; and the National Cancer Database currently are the largest registries available for study of lung cancer patients. Although well suited for epidemiologic studies, these registries do not contain detailed patient-level clinical, treatment, or outcome variables. These deficiencies present a major barrier to sophisticated comparative effectiveness research owing to bias and confounding. For example, none of these databases capture data on lung function, patient performance status, body mass index, and smoking history in addition to other important variables. Conversely, the STS GTSD captures unique



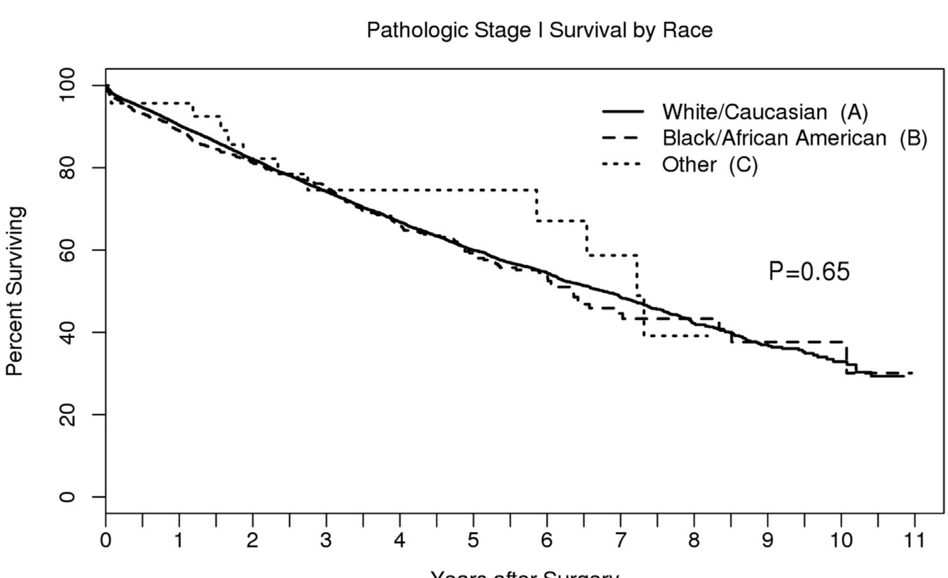
No. of Patients / Percent Survival

Female	9262	7037/93.0	5167/86.2	3605/79.0	2409/71.6	1513/65.4	858/59.8	450/54.2	221/48.4	76/43.4	32/40.0
Male	8134	5861/87.0	4145/77.1	2896/68.9	1918/61.1	1199/53.4	708/47.8	378/42.0	180/35.8	68/30.7	23/27.0

Fig 2. Overall survival after surgical resection of pathologic stage I lung cancer according to sex. Female patients are indicated by the solid line; male patients are indicated by the dashed line.

demographic and clinical data, surgical procedures, and complications.

Limitations to the STS-CMS linked dataset are acknowledged. As mentioned, this dataset is applicable to Medicare beneficiaries 65 years or older. Lung cancer surgery outcomes reported in the STS GTSD have also been shown to be superior to those reported from national administrative databases [13]. Such issues may limit the generalizability of the STS-CMS data. Further, the number of variables and definition of certain variables in the STS GTSD have changed over time with different versions of the data collection instrument. This is particularly true for tumor staging, as the granularity of the T descriptors captured has become more specific with time. Additionally, missing data are an important consideration for certain variables, such as clinical stage (10.7% missing). Finally, although the STS GTSD captures detailed clinical data, studies produced from the

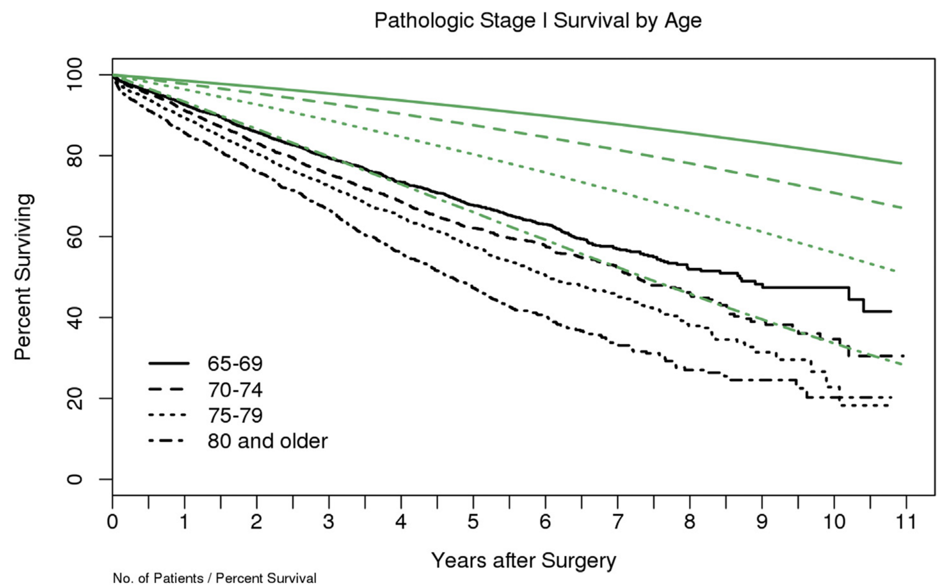


No. of Patients / Percent Survival

A	15632	11585/90.3	8320/82.0	5744/74.2	3806/66.8	2402/59.9	1389/54.4	735/48.6	354/42.2	127/37.0	49/32.9
B	959	709/89.0	502/81.3	364/75.3	222/66.3	141/59.3	78/53.8	35/44.6	23/43.3	8/37.6	5/37.6
C	46	31/95.7	24/82.2	18/74.5	16/74.5	12/74.5	9/67.1	7/58.7	2/39.1		

Fig 3. Overall survival after surgical resection of pathologic stage I lung cancer according to race. White/Caucasian is indicated by the solid line; black/African American is indicated by the dashed line; and other is indicated by the dotted line.

Fig 4. Overall survival after surgical resection of pathologic stage I lung cancer according to age group, with comparisons to age-expected survival (shown in green). Patients 65–69 years are indicated by the solid line; patients 70–74 years are indicated by the dashed line; patients 75–79 years are indicated by the dotted line; and patients 80 years and older are indicated by the dash-dot line.



No. of Patients / Percent Survival

65-69	4849	3663/92.6	2650/85.9	1859/79.5	1237/73.5	800/67.7	461/63.1	255/56.9	128/51.9	59/48.2	21/47.4
70-74	5202	3914/91.2	2801/83.1	1950/75.4	1346/68.6	860/62.1	520/57.7	282/52.3	142/46.2	50/39.0	22/34.7
75-79	4258	3140/89.4	2270/80.5	1612/72.6	1076/64.8	663/57.5	379/50.3	194/45.1	86/37.9	19/31.4	5/22.8
80 and older	3098	2191/85.6	1600/75.9	1088/66.6	676/55.8	396/47.2	207/40.3	97/33.4	45/27.0	16/24.5	7/20.2

STS-CMS linked dataset will be subject to unmeasured confounding, as are all retrospective analyses.

In conclusion, external linkage to CMS data enhances the value of the STS GTSD by providing a platform for longitudinal research initiatives in lung cancer surgery. Future studies will develop risk models for long-term survival after lung cancer surgery, examine the comparative effectiveness of different lung cancer resection strategies, and explore economic and resource use outcomes.

This project was supported by grant number R01 HS022279 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

References

1. Wright CD, Edwards FH, Society of Thoracic Surgeons General Thoracic Surgery Database Task Force; Society of Thoracic Surgeons Workforce on National Databases. The Society of Thoracic Surgeons General Thoracic Surgeons General Thoracic Surgery Database. Ann Thorac Surg 2007;83:893–4.
2. Jacobs JP, Edwards FH, Shahian DM, et al. Successful linking of The Society of Thoracic Surgeons Adult Cardiac Surgery Database to Centers for Medicare and Medicaid Services Medicare data. Ann Thorac Surg 2010;90:1150–7.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65:5–29.
4. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. Am Heart J 2009;157:971–82.
5. Magee MJ, Wright CD, McDonald D, Fernandez FG, Kozower BD. External validation of the society of thoracic surgeons general thoracic surgery database. Ann Thorac Surg 2013;96:1734–9.
6. National Center for Health Statistics. Life Tables. Available at http://www.cdc.gov/nchs/products/life_tables.htm. Accessed December 7, 2015.
7. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr 1961;6:101–21.
8. Shahian DM, O'Brien SM, Sheng S, et al. Predictors of long-term survival after coronary artery bypass grafting surgery: results from The Society of Thoracic Surgeons Adult Cardiac Surgery Database (The ASCERT Study). Circulation 2012;125:1491–500.
9. Weintraub WS, Grau-Sepulveda MV, Weiss JM, et al. Comparative effectiveness of revascularization strategies. N Engl J Med 2012;366:1467–76.
10. Pezzi CM, Mallin K, Mendez AS, Greer Gay E, Putnam JB Jr. Ninety-day mortality after resection for lung cancer is nearly double 30-day mortality. J Thorac Cardiovasc Surg 2014;148:2269–77.
11. Goldstraw P, Crowley J, Chansky K, et al, International Association for the Study of Lung Cancer International Staging Committee; Participating Institutions. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. J Thorac Oncol 2007;2:706–14.
12. 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.
13. Khullar OV, Gillespie T, Nickleach DC, et al. Socioeconomic risk factors for long-term mortality after pulmonary resection for lung cancer: an analysis of more than 90,000 patients from the National Cancer Data Base. J Am Coll Surg 2015;220:156–68.e4.

DISCUSSION

DR BENJAMIN D. KOZOWER (Charlottesville, VA): Thank you, Felix. I want to congratulate you and your colleagues at the STS [Society of Thoracic Surgeons] and also the DCRI [Duke Clinical Research Institute] for this outstanding work, and also to congratulate you for the R01 from AHRQ [Agency for Healthcare Research and Quality] that has funded this work. I do not think everyone realizes how much work goes into this; to combine two databases as complex as these are is extraordinary.

I think both of us agree that we in the surgical community and those working with the STS need to do a better job in assessing the true quality of care that we provide; and so to do this, we have to get beyond the traditional outcomes at 30 days.

So I have two questions for you. Because of age, HMO [health maintenance organization] status, and other factors, you were only able to match 43% of those patients in the GTSD [General Thoracic Surgery Database] to CMS [Centers for Medicare and Medicaid Services]. So how confident are you that this process provides a truly representative sample of long-term follow-up?

DR FERNANDEZ: Thanks, Ben. Thanks first for your comment, and I cannot emphasize enough how important the collaborative efforts were of all the coinvestigators, the STS, and the DCRI to make this possible.

The first comment I will make is we are restricted to linking to CMS because in 2011, the government restricted access to the Social Security Death Index, prohibiting long-term vital status follow-up on the entire cohort of patients. So the first restriction we have is to patients 65 and older.

So if you look at the under 65 group, they had more advanced stage disease. They had more extensive resections. They had lower operative mortality, and we also observed a slightly higher percentage of African Americans. So there are some differences between the under 65 and over 65 groups. So let us look at the 65 and over group.

The mean age of lung cancer resections in the STS database is 66 years; and the 65 and over group makes up 61% of the sample of lung cancer resections in the STS database.

When we compared this to the patients that were linked and not linked, we found fairly similar demographic characteristics. There is a slightly higher percentage of African Americans in the group, and there are some differences in the payer mix, and we cannot follow up patients with HMO and commercial insurance; and then there is a slightly higher operative mortality, within half a percentage point. These are things that need to be explored.

This, at present, is the best mechanism to provide long-term survival for lung cancer patients in the database. As you know and are leading, we are going to start prospectively doing this in the General Thoracic Surgery Database, but this data is a few years away.

We need to explore multiple avenues for obtaining long-term follow-up on our patients. One benefit that this Medicare data has is in addition to vital status, it gives us cost and resource use data, which is very important and very unique.

DR KOZOWER: Great, thanks. My second question has to do with data quality. We take great pride in the STS, and we talk about how we are externally audited, and the data is very valid, but you brought up a big red flag.

So we tout that the STS outperforms national benchmarks with low mortality and major complications, but you showed a pretty significant difference between perioperative mortality between the STS and the CMS; I think it is about half a percentage point. But when the outcome is so low, that is very significant.

So I wanted to know your thoughts about that. Do you know what types of patients those are? Do they have more in-hospital complications and prolonged length of stays, perhaps in a medical ICU [intensive care unit] or anything like that?

And the other is, what do we have to do as part of a GTSD task force to help improve that?

DR FERNANDEZ: Right. So the difference, when you add the CMS data, the mortality increases from 2.1% to 2.6% for lung cancer resections, and we have yet to drill down on that subset. It is something we need to do.

There are two reasons for this higher observed mortality, and we think the major reason is that the CMS data allows for more complete data capture, such as patients who may have died after they went home or maybe they were transferred to another facility or another medical service and became a nontraditional pathway patient, and were lost to follow-up.

The other thing that has to be considered is selective underreporting. Fortunately, our audits, which have been ongoing for several years now, have not uncovered any intentional gaming of the system.

So what do we do about it? As you know, this is a big point of emphasis for the database, for all three databases starting this year, to decrease the missing status or unknown status, for the vital status data field for patients at 30 days or hospital discharge.

I think how we do this is, number one, through education of the data managers. I think taking this data back to them at the AQO [Advances in Quality Outcomes] meeting will be of significant value because these are very dedicated people, and seeing this gap will increase their efforts.

We need to demand a certain standard of quality in the harvest data submitted to DCRI. As you know, there are going to be restrictions placed on the amount of missing data, in particular for that data field (vital status), that can be sent, so that will bounce back to you. And I think verification of life status with other data sources such as this one can help out in the process.

DR KOZOWER: Great, thank you very much.

DR VARUN PURI (St. Louis, MO): That was a fantastic paper. I really enjoyed your presentation. I have one comment and one quick question.

Your paper points us in the direction of what a utopia for an administrative database researcher or an outcomes researcher would be. That would be a single-payer system where everyone is under the same umbrella in terms of care and follow-up, so one would not have to go through the lengths of effort and great deals of difficulty that you encountered in trying to assemble this data to present it to us.

My quick question to you is you showed that there was not that much separation between the clinical and the pathologic stage I disease long-term outcomes. They were fairly similar with both of them showing, say, the 60% to 65%, 5-year survival in the clinical stage, and just about 60% or so in the pathologic stage. This was something which I found somewhat different in your data analysis than what we have seen from previous analyses. Would you like to comment about that?

DR FERNANDEZ: We observed the same thing, and we have not drilled down on upstaging yet and how this changes. And I think

an important other thing to note, the T descriptors have changed in the database. And I put this slide up here to illustrate what the T descriptors look like in the earlier versions of the database from 2002 to 2008, and how it has looked from 2009 to 2012, which is the last year we have linkage for.

So you can see that there is much more granularity in how we can stratify tumors according to the T descriptors, so that may have something to do with it. But it is 60% versus 56% for stage I disease. The separation is bigger than the IASLC [International Association for the Study of Lung Cancer] data. We need to explore that further.

DR SUDISH MURTHY (Cleveland, OH): Just a few brief comments, Felix. Very nice paper and presentation. This is a word of caution for the audience as they digest this material. The survival of patients with stage IV disease in this study is approximately 30%, and the audience needs to appreciate that patients were identified by these investigators through the STS database. This database tracks surgical patients. Consequently, stage IV disease in this study is either a solitary brain met[astasis] or other oligometastasis or synchronous primary tumors that were called stage IV disease by the coders. This is not the typical survival, nor patient population, found in traditional stage IV disease cohorts. I think you have to be careful to highlight this when you are putting this information out there.

DR FERNANDEZ: I could not agree more with your comments. These are highly selected patients, being selected from the best centers in the world, being operated on by the best surgeons.

DR LESLIE KOHMAN (Syracuse, NY): Just a question. What are the prospects of getting the Social Security Death Index unlocked?

DR FERNANDEZ: Well, there have been STS efforts for some time to, again, have access to the Social Security Death Index, and I am not privy to the latest updates.

As of yet, we do not have access to this data. Other medical specialties have been advocating for this as well.

DR PURI: Let me make one last comment. Just to answer the Social Security Death Index issue, I think the problem has been that the whole effort is underfunded currently in Washington, so it is about 9 to 12 months behind in terms of what is being reported to the SSDI. That is why they are not letting any organizations access this, because it is inaccurate currently, and that is the problem.