



Published in final edited form as:

J Thorac Cardiovasc Surg. 2016 March ; 151(3): 626–628. doi:10.1016/j.jtcvs.2015.08.110.

Large Clinical Databases for the Study of Lung Cancer: Making Up for the Failure of Randomized Trials

Chi-Fu Jeffrey Yang, MD¹, Matthew Hartwig, MD¹, Thomas A D'Amico, MD¹, and Mark F Berry, MD, MHS²

¹Department of Surgery, Duke University Medical Center, 3496 DUMC, Durham, North Carolina

²Department of Cardiothoracic Surgery, Stanford University Medical Center, 300 Pasteur Drive, Falk Building 2nd Floor Stanford, CA 94305-5407

Keywords

Lung cancer surgery; clinical database

Introduction

Although randomized controlled trials (RCT) represent the optimal study design for assessing the efficacy of clinical interventions, the practice of thoracic surgery has historically been significantly influenced by case series and single-institution observational cohort studies¹ due to the well-described difficulties of performing RCTs.² Researchers in the United States and Europe are increasingly utilizing large-scale population-based administrative databases³⁻⁵ and clinical registries⁶⁻¹⁰ to perform retrospective cohort studies to answer questions regarding the efficacy, cost and complications of interventions, to study quality improvement, to assess the treatment of rare conditions, and to evaluate national practice patterns. These studies can significantly enhance quantitative evidence regarding lung cancer treatment when performed by investigators who have clinical expertise in lung cancer with a thorough understanding of both the advantages and limitations of the database being utilized, through collaboration with experts in advanced statistical techniques.

Advantages of Analyzing Large Clinical Databases

Well-performed RCTs that minimize bias in patient selection provide the highest grade evidence to guide clinical practice. Although thoracic surgeons have performed extremely influential RCTs in the past,¹¹ many barriers to the use of RCT have generally limited the use of RCTs to investigate surgical interventions for lung cancer, including inadequate equipoise amongst surgeons and patients, complex institutional administrative requirements for RCT approval, difficulty in obtaining adequate, and inadequate research infrastructure in many programs.² Moreover, clinical trials involving thoracic surgical patients are commonly

Corresponding Author: Thomas A. D'Amico MD, Gary Hock Endowed Professor of Surgery, Chief, Section of General Thoracic Surgery, Duke University Medical Center, DUMC Box 3496, Duke South, White Zone, Room 3589, Durham, North Carolina 27710, Phone:(919) 681-0491, Fax: (919) 684-8508, thomas.damico@duke.edu.

M.F.B. has no disclosures to report.

terminated early due to poor accrual.¹²⁻¹⁴ As a result, most treatment guidelines for lung cancer surgery are based on experts' interpretations of data from case series and single-institution observational cohort studies.¹ The evidence from these study designs are more susceptible to bias and have lower “grades” than RCTs.

Well-done studies of large-scale population-based datasets can bridge the gaps in evidence that result from the dearth of data from RCTs. A major strength of clinical database studies is that their sample sizes are typically orders of magnitude larger than those of randomized controlled trials and single institution case series. For example, the National Cancer Data Base (NCDB) receives over one million cancer case reports annually⁶, and the NCDB contains data on more than 1.5 million patients with non-small cell lung cancer (NSCLC).¹⁵ The Medicare claims database has data on more than 45 million patients.¹⁶ The size of the available patient populations studied provide unprecedented statistical power to researchers and enables the performance clinically meaningful subset analyses that would be difficult to do with RCTs. For example, in the Lung Cancer Study Group (LCSG) randomized trial of lobectomy versus limited resection for T1 N0 NSCLC, investigators had found that limited resection was associated with increased locoregional recurrence for patients with tumors up to 3 cm in size but were unable to make any conclusions for smaller tumors due to lack of statistical power.¹¹ Two well-designed ongoing randomized trials – the Cancer and Leukemia Group B (CALGB) 140503 and the Japanese Clinical Oncology Group (JCOG) 0802 studies^{17, 18} – are focusing on tumors smaller than 2 cm, but may not have sufficient power to evaluate smaller tumors or compare differences between wedge vs segmentectomy vs lobectomy due to sample size and patient accrual issues. With the statistical power that comes from large clinical datasets, investigators have been able to evaluate the impact of lobectomy vs limited resection on survival for patients with smaller tumors less than 2 cm^{19, 20} and less than 1 cm²¹ as well as evaluate the differences in long-term outcomes between segmentectomy vs lobectomy,²² wedge vs segmentectomy²³ and wedge vs anatomic resection.²⁴

Considering that the highest level of evidence is given when multiple RCTs are available for a clinical situation, studies of large clinical registries or administrative claims databases can perform the important function of confirming results from both RCTs as well as smaller single-institution studies. For example, a study of thoracoscopic lobectomy at our institution demonstrated a lower morbidity compared to thoracotomy.²⁵ A subsequent national study using the Society of Thoracic Surgeons (STS) Database found similar results.²⁶ Although both studies were propensity-matched analyses and although the single-institution study was of a prospectively-collected database with excellent follow-up, the multi-institutional STS database provided a sample size six times larger than the single-institution study and was far more influential despite having limitations of voluntary participation, limited auditing of submitted data, and lack of follow-up beyond 30-days.

In addition, analyses of large-scale datasets have helped provide insight into the surgical management of topics that can be challenging to study, such as stage IIIA(N2) NSCLC treatment. Although a randomized clinical trial provided important data on the role of surgery for N2 disease, the trial's results also raised several questions²⁷ which have yet to be definitively answered with another randomized trial. Population-based datasets allow further

investigation in this topic whereas analysis of single-institution datasets are significantly limited due to both institutional biases and small sample sizes associated with the relatively rarity of this disease sub-stage.

Because analysis of large databases has data from real-world clinical practice, their results can be more generalizable than specialized single-institution studies or even a RCT, whose cohort may be healthier than the typical lung cancer patient population. In addition, large clinical databases facilitate development of risk-assessment tools (e.g. risk calculators), drive quality improvement and clinical governance, and help facilitate ongoing review of disease incidence, disease mortality, volume-outcome relationships and national trends in the use of procedures and disparities in health care.²⁸ Analyses from large databases can be used to screen various topics related to lung cancer, perhaps to identify which topics would be most appropriate for the investment of resources to conduct a large and expensive randomized trial.

Understanding and Overcoming the Limitations of Using Large Databases

All large-scale population-based databases have inherent limitations. Briefly, the STS and American College of Surgeons National Surgical Quality Improvement Program (NSQIP) databases are clinical registries with detailed co-morbidity and postoperative data but are limited in that they only have short-term outcomes.^{7, 9} Both the Surveillance, Epidemiology and End Results (SEER) Program and NCDB are cancer registries that have long-term survival data but are limited in that they do not have data on recurrence-free survival.^{6, 10} While the NCDB includes patient composite co-morbidity scores, it does not have detailed co-morbidity data and the SEER simply does not have any co-morbidity information.^{6, 10} The Medicare claims database and Nationwide Inpatient Sample (NIS) are administrative claims databases that rely on billing codes which may suffer from inaccuracies and variabilities in the coding; further Medicare data does not include patients younger than 65 years.^{4, 5} It is critically important that thoracic surgeons with clinical expertise in lung cancer be involved in the design of studies that use these datasets so that the limitations are both minimized and understood so that results can have important clinical implications. It is also imperative that thoracic surgeons foster collaboration with both biostatisticians and epidemiologists during these studies to achieve the highest possible impact.

Conclusion

Analyses of large-scale multi-institutional datasets are being increasingly performed by researchers to study lung cancer. These studies can provide important understanding of treatment at a population level and in many situations can enhance quantitative evidence regarding prognosis, efficacy of interventions, and disparities in treatment. Although these studies have limitations and cannot replace the gold standard of RCTs, appropriate use of these valuable datasets can enhance current evidence and also help direct future research endeavors.

Acknowledgments

Disclosures: This work was supported by the NIH funded Cardiothoracic Surgery Trials Network (M.G.H), 5U01HL088953-05 and by the American College of Surgeons Resident Research Scholarship (C.J.Y.). One of the authors (T.A.D.) serves as a consultant for Scanlan International, Inc.

References

1. Lee JS, Urschel DM, Urschel JD. Is general thoracic surgical practice evidence based? *The Annals of thoracic surgery*. 2000; 70:429–431. [PubMed: 10969657]
2. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ (Clinical research ed)*. 2002; 324:1448–1451.
3. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg*. 2012; 256:973–981. [PubMed: 23095667]
4. Houchens R, E A. 2012 National Inpatient Sample (NIS) Comparison Report. *HCUP Methods Series Report*. 2015 # 2015-04.
5. GF R. Administrative and claims records as sources of health care cost data. *Med Care*. 2009; 47:S51–55. [PubMed: 19536019]
6. Bilimoria KYSA, Winchester DP, Ko CY. The National Cancer Data Base: a powerful initiative to improve cancer care in the United States 2008. 2008; 15:683–690.
7. Shahian DMJJ, J J, Edwards FH, Brennan JM, Dokholyan RS, Prager RL, Wright CD, Peterson ED, McDonald DE, Grover FL. The society of thoracic surgeons national database. *Heart*. 2013; 99:1494–1501. [PubMed: 23335498]
8. Fernandez FG, F P, Kozower BD, Salati M, Wright CD, Brunelli A. The Society of Thoracic Surgeons and the European Society of Thoracic Surgeons general thoracic surgery databases: joint standardization of variable definitions and terminology. *The Annals of thoracic surgery*. 2015; 99:368–376. [PubMed: 25555970]
9. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005; 138:837–843. [PubMed: 16291383]
10. Yu JBGC, Wilson LD, Smith BD. NCI SEER public-use data: applications and limitations in oncology research. *Oncology (Williston Park)*. 2009; 23:288–295. [PubMed: 19418830]
11. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. *Ann Thorac Surg*. 1995; 60:615–622. discussion 622-613. [PubMed: 7677489]
12. Schroen AT, Petroni GR, Wang H, et al. Preliminary evaluation of factors associated with premature trial closure and feasibility of accrual benchmarks in phase III oncology trials. *Clin Trials*. 2010; 7:312–321. [PubMed: 20595245]
13. Baggstrom MQ, Waqar SN, Sezhiyan AK, et al. Barriers to enrollment in non-small cell lung cancer therapeutic clinical trials. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*. 2011; 6:98–102.
14. Martins RG, D'Amico TA, Loo BW Jr, et al. The management of patients with stage IIIA non-small cell lung cancer with N2 mediastinal node involvement. *Journal of the National Comprehensive Cancer Network: JNCCN*. 2012; 10:599–613. [PubMed: 22570291]
15. [Accessed on 8/27/2015] NCDB Public Benchmark Reports. Cases Diagnosed 2003-2013. <http://oliver.facs.org/BMPub/index.cfm>
16. Pugely AJ, Martin CT, Harwood J, Ong KL, Bozic KJ, Callaghan JJ. Database and Registry Research in Orthopaedic Surgery: Part I: Claims-Based Data. *J Bone Joint Surg Am*. 2015; 97:1278–1287. [PubMed: 26246263]
17. Nakamura K, Saji H, Nakajima R, et al. A phase III randomized trial of lobectomy versus limited resection for small-sized peripheral non-small cell lung cancer (JCOG0802/WJOG4607L). *Japanese journal of clinical oncology*. 2010; 40:271–274. [PubMed: 19933688]
18. [Last accessed on November 22, 2013] CALGB 140503. A phase III Randomized trial of lobectomy versus sublobar resection for small (\leq 2cm) peripheral non-small cell lung cancer. http://www.calgb.org/Public/meetings/presentations/2007/cra_ws/03-140501-Altorki062007.pdf

19. Wisnivesky JP, Henschke CI, Swanson S, et al. Limited resection for the treatment of patients with stage IA lung cancer. *Annals of surgery*. 2010; 251:550–554. [PubMed: 20160639]
20. Speicher P, Gu L, Gulack BC, Wang X, D'Amico TA, Hartwig MG, Berry MF. Sublobar Resection for Clinical Stage IA Non-small Cell Lung Cancer in the United States. *Clinical lung cancer*.
21. Kates M, Swanson S, Wisnivesky JP. Survival following lobectomy and limited resection for the treatment of stage I non-small cell lung cancer<=1 cm in size: a review of SEER data. *Chest*. 2011; 139:491–496. [PubMed: 20576736]
22. Whitson BA, Groth SS, Andrade RS, Maddaus MA, Habermann EB, D'Cunha J. Survival after lobectomy versus segmentectomy for stage I non-small cell lung cancer: a population-based analysis. *The Annals of thoracic surgery*. 2011; 92:1943–1950. [PubMed: 21962268]
23. Yang CF, C D, Gulack BC, Speicher PJ, Onaitis MW, Tong BC, D'Amico TA, Harpole DH, Berry MF, Hartwig MG. Wedge Resection vs Segmentectomy for Patients with T1A N0 Nonsmall Cell Lung Cancer. 16th World Conference on Lung Cancer. 2015
24. Linden PA, D'Amico TA, Perry Y, et al. Quantifying the safety benefits of wedge resection: a society of thoracic surgery database propensity-matched analysis. *Ann Thorac Surg*. 2014; 98:1705–1711. discussion 1711-1702. [PubMed: 25201723]
25. Villamizar NR, Darrabie MD, Burfeind WR, et al. Thoracoscopic lobectomy is associated with lower morbidity compared with thoracotomy. *The Journal of thoracic and cardiovascular surgery*. 2009; 138:419–425. [PubMed: 19619789]
26. Paul S, Altorki NK, Sheng S, et al. Thoracoscopic lobectomy is associated with lower morbidity than open lobectomy: a propensity-matched analysis from the STS database. *The Journal of thoracic and cardiovascular surgery*. 2010; 139:366–378. [PubMed: 20106398]
27. Albain KS, Swann RS, Rusch VW, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. *Lancet*. 2009; 374:379–386. [PubMed: 19632716]
28. Saxena A, Newcomb AE, Dhurandhar V, Bannon PG. Application of Clinical Databases to Contemporary Cardiac Surgery Practice: Where are we now? *Heart Lung Circ*. 2015

Central Message

The increasing use of large, multi-institutional datasets by researchers to study lung cancer has improved the understanding of treatments at a population level and in many situations can enhance quantitative evidence regarding prognosis, comparative efficacy of various interventions, and disparities in treatment.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript