Journal of Gastrointestinal Surgery (2021) 25:2989–2996
https://doi.org/10.1007/s11605-021-05094-y

REVIEW ARTICLE

Strengths and Limitations of Registries in Surgical Oncology Research

Sivesh K Kamarajah[1,2,3] · Hari Nathan[4,5]

Received: 2 March 2021 / Accepted: 11 July 2021 / Published online: 10 September 2021

Abstract

Over the past two decades, there has been a dramatic increase in studies based on large multi-institutional tumor registries. Applications of such databases span various research themes including epidemiology, oncology, surgical techniques, perioperative outcomes, and prognosis. Although these databases are acquired relatively easily, offer larger sample sizes and improved generalizability compared with institutional data, acknowledging limitations within analysis and cautious interpretation of data is important. Questionable conclusions can result when insufficient attention is paid to issues such as data quality and depth, potential sources of bias and missing data. This article reviews research themes and important limitations of these databases. The contemporary reporting of these issues in the literature and an increased awareness among surgical oncologists of potential applications and limitations will ensure that studies in the surgical oncology literature achieve high standards of methodological quality and clinical utility.

Keywords NCDB; · SEER; · Surgical oncology; · Gastrointestinal surgery; · Health services research

Introduction

Over the past two decades, there has been a dramatic increase in studies based on large databases such as administrative claims data and large multi-institutional tumor registries (Fig. 1). In the field of surgical oncology, the Surveillance, Epidemiology, and End Results (SEER) registry[1] maintained by the National Cancer Institute (NCI) has been used extensively, both alone and in conjunction with administrative data from the Medicare program,[2] to understand epidemiology, treatment, and prognosis. Additionally, the National Cancer Database (NCDB) maintained by the American College of Surgeons Commission on Cancer has become a popular source of data on surgical and multimodality treatment and outcomes at cancer-focused hospitals. These data have been used to explore various themes including epidemiology, oncology, surgical techniques, perioperative outcomes, and prognosis. The relative ease with which these data can be acquired, the large available sample sizes, their improved generalizability compared with institutional data and reliable long-term follow-up collectively make them attractive to researchers. Furthermore, the availability of long-term follow-up offers an attractive option to study long-term survival of cancer patients.

Although registry analyses have an important role in surgical research, the limitations of these data and the statistical methods used to analyze the data often do not receive sufficient attention in the surgical literature. A strong understanding of the strengths and weaknesses of these databases is necessary to allow surgeons to perform high-quality research and critically appraise the literature. This review seeks to provide an overview of common research themes with some well-described limitations and caveats specific to the analysis and interpretation of these data. Although examples from SEER and NCDB are specifically cited, the concepts reviewed are broadly generalizable to secondary data analyses using other databases.

✉ Hari Nathan
drnathan@umich.edu

[1] Department of Surgery, Queen Elizabeth Hospital Birmingham, Birmingham, UK

[2] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

[3] Department of HPB and Transplant Surgery, The Freeman Hospital, Newcastle upon Tyne, Tyne and Wear, UK

[4] Department of Surgery, University of Michigan, Ann Arbor, MI, USA

[5] University of Michigan, 2210A Taubman Health Care Center, 1500 E Medical Center Dr, SPC 5343, Ann Arbor, MI 48109-5343, USA

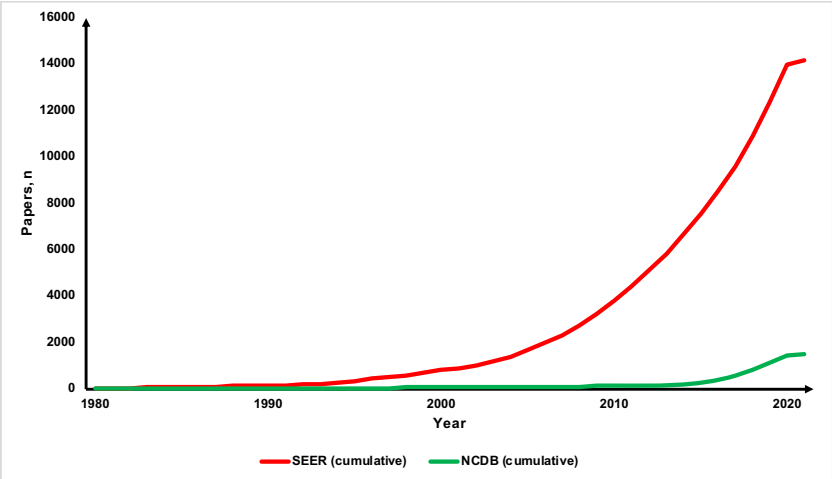

Fig. 1 Trends in use of SEER and NCDB over the past few decades

Background

The Surveillance, Epidemiology, and End Results (SEER) database is maintained by the National Cancer Institute (NCI). The SEER database has grown to include 21 cancer registries, representing 28% of the US population. As compared to the general US population, the SEER population is slightly more urban and has a slightly higher percentage of foreign-born individuals. Available data include patient demographics, tumor characteristics (e.g., histology, grade, stage), treatment data (e.g., surgery, radiation), and survival. Some data elements (e.g., American Joint Commission on Cancer (AJCC) staging, details of surgical therapy, tumor size, lymph node involvement) are consistently available only in more recent time periods. The NCI staff work with the North American Association of Central Cancer Registries (NAACCR) External Web Site Policy and a number of collaborating organizations to guide all state registries to achieve data content and compatibility acceptable for pooling data and improving national estimates.

The NCDB is a joint project of the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society.[3,4] The NCDB gathers information from approximately 1,500 CoC-accredited hospitals and includes >70 % of all newly diagnosed malignancies in the USA. Because of its reliance on CoC hospitals for data, NCDB is not, strictly speaking, a population-based database. Similar to SEER, NCDB contains specific details about patient demographics, tumor characteristics, treatment, and survival. NCDB also includes information on patient comorbidities, facility type and location, perioperative outcomes (e.g., length of stay, readmission), receipt of chemotherapy, and resection margins. In addition, Through linkage to tertiary data sources, area-based socioeconomic status and estimates of travel distance/time to the reporting hospital can be obtained based on each patient's home zip code at the time of diagnosis. In addition, hospital type and structural characteristics can be determined through linkage with the CoC's facility information profile system (FIPS) file of self-reported hospital structural features, resources, and services related to oncology.

Research Themes

Epidemiology

Assessment of cancer incidence has been a major strength of SEER, providing population level data to allow age-adjusted analysis. These include incidence and mortality rates across different cancers over time. A notable example is the annual cancer statistics report released by the American Cancer Society evaluating trends in cancer incidence and mortality across different cancers over time.[5–7] The large numbers allow study of specific demographic subgroups and uncommon disease types and histologies. Registry analyses also help provide benchmarking data for quality improvement and assessment of disparities in access, quality, and outcomes.

Surgical Oncology

The NCDB allows for multimodal treatment to be characterized, while this is less so for SEER as they lack information on type of oncological therapy (e.g., chemotherapy, radiotherapy). For instance, NCDB provides type (single-agent vs. multi-agent), sequence (neoadjuvant, intraoperative, adjuvant), and duration of chemotherapy, which potentially allows more nuanced analysis compared to the SEER database, where these data are not available. However, it still remains difficult to separate chemoradiation and chemotherapy in the NCDB.

The NCDB codes all events relative to the date of diagnosis (the first date a statement of diagnosis is made by a physician, which may be based on results of imaging or confirmation via histologic testing). Therefore, the timing of various treatments

or events (e.g., start of chemotherapy or surgery) relative to the date of diagnosis can be used to determine sequence.[8] More complicated strategies (e.g., chemotherapy or radiotherapy before and after surgery) can be determined using specific sequence variables in the NCDB. Recent studies have used timing data in the NCDB to assess the association of delaying adjuvant chemotherapy on survival of colon cancer[9] and lung cancer.[10] Further, NCDB allows assessment of national practices across the USA on use of systemic chemotherapy or radiotherapy (RT) in specific cancer types and evaluate its long-term benefit. For instance, a recent study highlighted a modest survival benefit of adjuvant RT in patients undergoing margin-negative pancreatic cancer resections using appropriate propensity matching techniques to account for treatment selection bias.[11]

Owing to the availability of some information on surgical approach (e.g., open or minimally invasive), researchers have used NCDB in assessing the association of minimally invasive techniques on short-term and long-term outcomes across a spread of different cancers.[12–15] These data are not available in SEER. Such data have allowed appraisal on the benefit of minimally invasive techniques beyond a few specialized centers. More recently, the adoption of robotic surgery has led to further analysis on the benefits of robotic over laparoscopic surgery in surgical oncology cancers.[16–19] However, it is important to note that current appraisals on minimally invasive techniques may be limited as surgery types (e.g., two-stage, three-stage, transhiatal esophagectomies) and reasons for conversion to open surgery are not available to provide granular analysis, resulting in some degree of bias in interpretation of results.

Health Services Research

Common themes within health services research exploring care of cancer patients include determinants of short-term outcomes (e.g., teaching status, center volume), determinants of long-term outcomes (e.g., time to surgery, timing of adjuvant treatment), hospital-based variation in treatment and outcomes, and disparities in treatment and outcomes. While center volume has been widely reported to be associated with short-term mortality,[20,21] data on its impact on long-term survival of cancer patients are less robust. Center volume data can be derived from the NCDB through available unique facility codes, but this is not the case for the SEER database (unless Medicare claims are used).[22] Studies have demonstrated improved survival in patients treated in high-volume centers for glioblastoma,[23] metastatic renal cell carcinoma,[24] and prostate cancers,[25] for example.

The impact of time to surgery on long-term survival has also been evaluated using these databases as a determinant of outcomes. Although some studies have demonstrated that a delay to surgery is associated with reduced survival[26–30] due to theoretical increased risk in delay and hence progression of tumor, some studies have reported the contrary. These data may be useful for clinicians for counselling patients on the impact on delay to surgery. However, cautious interpretation is required as such analysis may not consider residual bias such as from potential delays from referral or patients' receipt neoadjuvant therapy prior to surgery.

NCDB and SEER have been utilized to understand impact of hospital-level characteristics such as center volume and institutional type (e.g., academic vs non-academic). In particular, NCDB has been utilized to examine impact of center volume in short-term mortality following major cancer surgery, owing to availability of annual procedural volume. However, this is not the case for the SEER database. Studies from NCDB have demonstrated that high-volume centers were associated with significantly lower 30-day mortality rates compared to low-volume centers, despite similar complication rates.[31–33] These data implicate failure to rescue as a major determinant of variation in hospital mortality rates. In other work, centers designated as "academic" were more likely to have better surgical outcomes than "non-academic" centers.[23] It is difficult to assess the significance of such findings, as programs with academic research are highly heterogeneous. Whenever possible, more detailed and objective hospital characteristics should be used. Although details of these observations are not available within both databases, they provide impetus for future research to identify factors that may explain these findings.[21,34,35]

Furthermore, underutilization of surgery in early stage cancers has also been a popular topic within both databases to highlight disparities in allocation to curative surgery. Several studies highlight variation in trends of curative resection for early (stage I) cancers,[36,37] for which resection is usually the gold standard treatment. These studies have three broad advantages: (i) evaluate and characterize the utilization of surgery for cancers; (ii) to identify factors predicting failure to undergo surgery for localized disease, and (iii) to evaluate the association of surgery with survival. Such studies help identify areas for improvement and target center- or population-level interventions to improve utilization of surgery across centers.

Perioperative Outcomes

Research questions around perioperative outcomes are particularly common with the NCDB (but not SEER) due to the availability of readmission, 30-day and 90-day mortality rates with the former. Recent studies[38–40] have evaluated readmissions and 30-day mortality in across different cancers in identifying factors associated with readmissions as these Interventions aimed at decreasing hospital readmissions should target these high-risk patients. These studies are limited by the fact that reasons for readmission are not captured

within the NCDB. This is important especially in the context of identifying areas for improvement within health systems.

Prognosis and Staging

In the theme of prognostication, SEER and NCDB have been commonly used in validation of the AJCC staging system across various cancers over the past decade. Such analyses have been useful in testing the relevance of new changes to these staging systems and assess performance in models.[41–44] Although these studies identify that current staging systems are poor at discriminating long-term survival in patients undergoing curative resection, it helps clinicians in providing important information to patients. It is important to note that these databases can only assess factors that are collected by the databases, which are usually just the factors that are already in the AJCC staging system

Owing to limitations of existing staging systems for cancers, both the SEER and NCDB registries have been also been utilized in developing prognostic models, risk scores or nomograms in guiding discussions with patients to guide treatment decisions that may influence long-term survival. For instance, a scoring system was developed to predict risk of lymph node metastases in patients with early T1 esophageal cancers to guide clinicians on decision between endoscopic or surgical resection.[45] The proposed scoring system seems to be useful in discriminating risk of nodal metastases in patients with T1a or T1b esophageal adenocarcinoma and may be useful in directing patients who received endoscopic resection to esophagectomy or careful follow-up. Similar studies have been performed in other cancers with NCDB or SEER.[46] However, applicability of these risk scores in other populations remain unknown as these risk scores lack external validation.

Finally, both NCDB and SEER are useful in identifying patient-level, tumor-level and treatment-level characteristics associated with good long-term prognosis. This remains possible with NCDB and SEER due to good long-term follow-up of these patients. A recent NCDB study in patients with pancreatic cancers identified only 4% (431/11,081) of patients were alive at 10 year and significant predictors were lymph node ratio, administration of adjuvant chemotherapy, and pathologic T stage.[47] These useful well-designed study will help clinicians in prognostication of patient following cancer resections. Wider applicability to a variety of different cancer types can aid decision making and planning of follow-up management.

Addressing Database Limitations

Generalizability and Missing Data

Although SEER and NCDB provide comprehensive evaluation across >1000 institutions across the USA, neither of these prospectively maintained registries are externally validated to ensure high case ascertainment and data accuracy. Lack of external validation is highlighted by the variation in missing data across data fields, an inherent issue widely recognized within large databases. As a general rule, data fields with >10% missing data threaten data validity due to risk of bias. Two options exists to account for missing data: (1) Listwise deletion or complete-case analysis[48]; or (2) multiple imputation.[49] The former is subjected to lower available sample size and consequently results in loss of power. Multiple imputation is one way to ensure the sample size is maintained to ensure reliable results can be drawn from powered studies. This is especially important in diseases with low incidence such as cholangiocarcinoma[50–53] or gallbladder cancers.[54,55] However, even for multiple imputation to be used, missing data should be missing at random, a condition that often cannot be empirically proven. Studies using administrative data should report the extent of missing data, use appropriate methods to account for missing data in analyses, and discuss their potential impact on inferences and conclusions.

Data Quality and Depth

Across cancer registries, there remain issues to depth of data provided, limiting granular analyses. For instance, NCDB contains data on margin status, surgical approach (e.g., open, laparoscopic, or robotic), and adjuvant regimens (e.g., chemotherapy, chemoradiotherapy). However, NCDB has some limitations in this regard. Firs, margin status may be biased by variation in institutional assessment because definitions for margin-negative or margin-positive resections are not well-defined. This is particularly important for esophagogastric cancers where details of longitudinal or circumferential margins are not specified in the database. Second, the Charlson-Deyo comorbidity score is provided in the NCDB but is captured in a summative way that does not convey fitness for surgery. Furthermore, data on performance status are not available within these datasets. Third, another notable difference exists between cancer types. For example, comprehensive genetic information is provided for breast and colorectal cancers, but not for hepatobiliary and esophagogastric cancers. Fourth, there is no information on disease-free or disease-specific survival. Finally, both databases also lack clinical detail beyond staging, such as extent of vascular involvement in pancreatic cancer or details of multivisceral involvement of solid tumors. In liver cancer, for example, location of a lesion adjacent to critical structures and extent of underlying liver disease can affect treatment options but are not conveyed in the data.

Table 1 Summary of advantages and disadvantages of the NCDB and SEER databases

	NCDB	SEER
Advantages	Captures ~70% of incident cancers in the US	Captures ~34% of cancers in the US currently
	Granular tumor specific and treatment data including staging, chemotherapy, radiation, surgery	Contains pathological staging information
	Tracks overall survival	Tracks overall and cancer-specific long-term survival
		Cause of death is available
Disadvantages	No smoking status	No smoking status
	Only comorbidity information is Charlson comorbidity index	No comorbidity data
	Only tracks first course of treatment	No chemotherapy data
	No recurrence information	Smaller sample than NCDB
	No cancer-specific long-term survival	

Bias

Although external validity of large databases can be threatened if the study population does not accurately represent the clinical population of interest, internal validity can be threatened by biases.[52] The three main types of widely recognized biases are information bias, selection bias, and confounding bias.[56] However, treatment selection bias remains pertinent to surgical oncology research. Non-random treatment assignment to oncological treatment introduces selection biases. For instance, allocation to surgery is more likely in younger, less co-morbid patients or those with early stage disease. To deal with this, multivariable regression modelling is utilized to produce adjusted estimates of the explanatory variable of interest.[57] Over the past few year, propensity score analysis[58] to account for non-random treatment assignment has become an increasingly popular method to address treatment selection bias. Despite this, correct application of propensity matching within surgical literature is lacking, warranting further judicious use to ensure accurate and reliable conclusions are drawn.[59] Furthermore, these biases may be overcome with the use of interaction and sensitivity analyses to support findings of the overall analysis as utilized in recent NCDB studies.[11,60]

Conclusion

Moving forwards, cancer registries will be useful to assess relevant questions pertinent to national clinical practice. Large database studies will be able provide a "real" world perspective on surgical practice and could aid in conducting well-designed randomized controlled trials (RCTs). There even have been initiatives to integrated the strengths of large database studies and RCTs by performing for example registry-based pragmatic RCTs.[61] In addition, large database studies are able to reflect on research questions regarding the efficacy of health policy, access to health care, and trends and geographic variation in practice patterns, as well as the treatment of rare disease or patients subgroups, which would be impossible or very strenuous by using RCTs. Moreover, large nationwide datasets, provide the tremendous opportunity to benchmark surgical outcomes and subsequently improve quality of care.

However, judicious use of these databases is also important, as analyses using these data are susceptible to bias (Table 1). Therefore, we have summarized a checklist to assist researchers in the appropriate use of these databases

Table 2 Summary checklist for research studies utilizing the NCDB and SEER databases

RESEARCH QUESTION
Can the question be reasonably answered using SEER or NCDB?
Are the inclusion and exclusion criteria clearly and correctly specified?
Have the intervention and/or comparator arms been appropriately defined?
MISSING DATA
What is the proportion of missing data in variables needed for the study?
How will data missingness be addressed?
How could data missingness impact my results?
DATA DEPTH
What are the known confounders that are not included in the database?
Do they threaten the validity of the analysis and conclusions?
Can their impact on the conclusions be estimated?
BIAS
Can threats to validity be addressed with additional analyses?
Can sources of bias and the size and direction of their impact on the conclusions be described?

(Table 2). Great emphasis on appropriate, well-designed studies and statistical methods are essential, but this does not obviate the inherent limitations of the data or their retrospective nature. Statistical methods should aim to explore and mitigate threats to the validity of the conclusions whenever possible. Researchers and reviews alike should focus not on the limitations of the data per se—there are many—but rather on how these limitations affect the inferences from the data and conclusions of the research.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Duggan, M. A.; Anderson, W. F.; Altekruse, S.; Penberthy, L.; Sherman, M. E., The Surveillance, Epidemiology, and End Results (SEER) Program and Pathology: Toward Strengthening the Critical Relationship. *Am J Surg Pathol* 2016, 40 (12), e94-e102.
2. Warren, J. L.; Klabunde, C. N.; Schrag, D.; Bach, P. B.; Riley, G. F., Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 2002, 40 (8 Suppl), IV-3-18.
3. Bilimoria, K. Y.; Bentrem, D. J.; Ko, C. Y.; Ritchey, J.; Stewart, A. K.; Winchester, D. P.; Talamonti, M. S., Validation of the 6th edition AJCC Pancreatic Cancer Staging System: report from the National Cancer Database. *Cancer* 2007, 110 (4), 738-44.
4. Merkow, R. P.; Rademaker, A. W.; Bilimoria, K. Y., Practical Guide to Surgical Data Sets: National Cancer Database (NCDB). *JAMA Surg* 2018.
5. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2019. *CA Cancer J Clin* 2019, 69 (1), 7-34.
6. DeSantis, C. E.; Miller, K. D.; Goding Sauer, A.; Jemal, A.; Siegel, R. L., Cancer statistics for African Americans, 2019. *CA Cancer J Clin* 2019, 69 (3), 211-233.
7. Miller, K. D.; Nogueira, L.; Mariotto, A. B.; Rowland, J. H.; Yabroff, K. R.; Alfano, C. M.; Jemal, A.; Kramer, J. L.; Siegel, R. L., Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin* 2019, 69 (5), 363-385.
8. Hanna, N.; Sun, M.; Meyer, C. P.; Nguyen, P. L.; Pal, S. K.; Chang, S. L.; de Velasco, G.; Trinh, Q. D.; Choueiri, T. K., Survival Analyses of Patients With Metastatic Renal Cancer Treated With Targeted Therapy With or Without Cytoreductive Nephrectomy: A National Cancer Data Base Study. *J Clin Oncol* 2016, 34 (27), 3267-75.
9. Sun, Z.; Adam, M. A.; Kim, J.; Nussbaum, D. P.; Benrashid, E.; Mantyh, C. R.; Migaly, J., Determining the Optimal Timing for Initiation of Adjuvant Chemotherapy After Resection for Stage II and III Colon Cancer. *Dis Colon Rectum* 2016, 59 (2), 87-93.
10. Salazar, M. C.; Rosen, J. E.; Wang, Z.; Arnold, B. N.; Thomas, D. C.; Herbst, R. S.; Kim, A. W.; Detterbeck, F. C.; Blasberg, J. D.; Boffa, D. J., Association of Delayed Adjuvant Chemotherapy With Survival After Lung Cancer Surgery. *JAMA Oncol* 2017, 3 (5), 610-619.
11. Kamarajah, S. K.; Sonnenday, C. J.; Cho, C. S.; Frankel, T. L.; Bednar, F.; Lawrence, T. S.; Nathan, H., Association of Adjuvant Radiotherapy With Survival After Margin-negative Resection of Pancreatic Ductal Adenocarcinoma: A Propensity-matched National Cancer Database (NCDB) Analysis. *Ann Surg* 2019.
12. Yang, C. J.; Kumar, A.; Klapper, J. A.; Hartwig, M. G.; Tong, B. C.; Harpole, D. H., Jr.; Berry, M. F.; D'Amico, T. A., A National Analysis of Long-term Survival Following Thoracoscopic Versus Open Lobectomy for Stage I Non-small-cell Lung Cancer. *Ann Surg* 2019, 269 (1), 163-171.
13. Skancke, M.; Schoolfield, C.; Umapathi, B.; Amdur, R.; Brody, F.; Obias, V., Minimally Invasive Surgery for Rectal Adenocarcinoma Shows Promising Outcomes Compared to Laparotomy, a National Cancer Database Observational Analysis. *J Laparoendosc Adv Surg Tech A* 2019, 29 (2), 218-224.
14. Gabriel, E.; Thirunavukarasu, P.; Attwood, K.; Nurkin, S. J., National disparities in minimally invasive surgery for pancreatic tumors. *Surg Endosc* 2017, 31 (1), 398-409.
15. Klein, G.; Wang, H.; Elshabrawy, A.; Nashawi, M.; Gourley, E.; Liss, M.; Kaushik, D.; Wu, S.; Rodriguez, R.; Mansour, A. M., Analyzing National Incidences and Predictors of Open Conversion During Minimally Invasive Partial Nephrectomy for cT1 Renal Masses. *J Endourol* 2020.
16. Hoehn, R. S.; Nassour, I.; Adam, M. A.; Winters, S.; Paniccia, A.; Zureikat, A. H., National Trends in Robotic Pancreas Surgery. *J Gastrointest Surg* 2020.
17. Thirunavukarasu, P.; Gabriel, E.; Attwood, K.; Kukar, M.; Hochwald, S. N.; Nurkin, S. J., Nationwide analysis of short-term surgical outcomes of minimally invasive esophagectomy for malignancy. *Int J Surg* 2016, 25, 69-75.
18. Arnold, B. N.; Thomas, D. C.; Narayan, R.; Blasberg, J. D.; Detterbeck, F. C.; Boffa, D. J.; Kim, A. W., Robotic-Assisted Lobectomies in the National Cancer Database. *J Am Coll Surg* 2018, 226 (6), 1052-1062 e15.
19. Raoof, M.; Nota, C.; Melstrom, L. G.; Warner, S. G.; Woo, Y.; Singh, G.; Fong, Y., Oncologic outcomes after robot-assisted versus laparoscopic distal pancreatectomy: Analysis of the National Cancer Database. *J Surg Oncol* 2018, 118 (4), 651-656.
20. Birkmeyer, J. D.; Siewers, A. E.; Finlayson, E. V.; Stukel, T. A.; Lucas, F. L.; Batista, I.; Welch, H. G.; Wennberg, D. E., Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002, 346 (15), 1128-37.
21. Ghaferi, A. A.; Birkmeyer, J. D.; Dimick, J. B., Hospital volume and failure to rescue with high-risk surgery. *Med Care* 2011, 49 (12), 1076-81.
22. Potosky, A. L.; Warren, J. L.; Riedel, E. R.; Klabunde, C. N.; Earle, C. C.; Begg, C. B., Measuring complications of cancer treatment using the SEER-Medicare data. *Med Care* 2002, 40 (8 Suppl), IV-62-8.
23. Zhu, P.; Du, X. L.; Zhu, J. J.; Esquenazi, Y., Improved survival of glioblastoma patients treated at academic and high-volume facilities: a hospital-based study from the National Cancer Database. *J Neurosurg* 2019, 132 (2), 491-502.
24. Joshi, S. S.; Handorf, E. A.; Zibelman, M.; Plimack, E. R.; Uzzo, R. G.; Kutikov, A.; Smaldone, M. C.; Geynisman, D. M., Treatment Facility Volume and Survival in Patients with Metastatic Renal Cell Carcinoma: A Registry-based Analysis. *Eur Urol* 2018, 74 (3), 387-393.
25. Chen, Y. W.; Mahal, B. A.; Muralidhar, V.; Nezolosky, M.; Beard, C. J.; Den, R. B.; Feng, F. Y.; Hoffman, K. E.; Martin, N. E.; Orio, P. F.; Nguyen, P. L., Association Between Treatment at a High-Volume Facility and Improved Survival for Radiation-Treated Men With High-Risk Prostate Cancer. *Int J Radiat Oncol Biol Phys* 2016, 94 (4), 683-90.

26. Kaltenmeier, C.; Shen, C.; Medich, D. S.; Geller, D. A.; Bartlett, D. L.; Tsung, A.; Tohme, S., Time to Surgery and Colon Cancer Survival in the United States. *Ann Surg* 2019.
27. Bleicher, R. J.; Ruth, K.; Sigurdson, E. R.; Beck, J. R.; Ross, E.; Wong, Y. N.; Patel, S. A.; Boraas, M.; Chang, E. I.; Topham, N. S.; Egleston, B. L., Time to Surgery and Breast Cancer Survival in the United States. *JAMA Oncol* 2016, 2 (3), 330-9.
28. Samson, P.; Patel, A.; Garrett, T.; Crabtree, T.; Kreisel, D.; Krupnick, A. S.; Patterson, G. A.; Broderick, S.; Meyers, B. F.; Puri, V., Effects of Delayed Surgical Resection on Short-Term and Long-Term Outcomes in Clinical Stage I Non-Small Cell Lung Cancer. *Ann Thorac Surg* 2015, 99 (6), 1906-12; discussion 1913.
29. Rygalski, C. J.; Zhao, S.; Eskander, A.; Zhan, K. Y.; Mroz, E. A.; Brock, G.; Silverman, D. A.; Blakaj, D.; Bonomi, M. R.; Carrau, R. L.; Old, M. O.; Rocco, J. W.; Seim, N. B.; Puram, S. V.; Kang, S. Y., Time to Surgery and Survival in Head and Neck Cancer. *Ann Surg Oncol* 2020.
30. Kompelli, A. R.; Li, H.; Neskey, D. M., Impact of Delay in Treatment Initiation on Overall Survival in Laryngeal Cancers. *Otolaryngol Head Neck Surg* 2019, 160 (4), 651-657.
31. Wong, S. L.; Revels, S. L.; Yin, H.; Stewart, A. K.; McVeigh, A.; Banerjee, M.; Birkmeyer, J. D., Variation in hospital mortality rates with inpatient cancer surgery. *Ann Surg* 2015, 261 (4), 632-6.
32. Wright, J. D.; Huang, Y.; Ananth, C. V.; Tergas, A. I.; Duffy, C.; Deutsch, I.; Burke, W. M.; Hou, J. Y.; Neugut, A. I.; Hershman, D. L., Influence of treatment center and hospital volume on survival for locally advanced cervical cancer. *Gynecol Oncol* 2015, 139 (3), 506-12.
33. Chapman, B. C.; Paniccia, A.; Hosokawa, P. W.; Henderson, W. G.; Overbey, D. M.; Messersmith, W.; McCarter, M. D.; Gleisner, A.; Edil, B. H.; Schulick, R. D.; Gajdos, C., Impact of Facility Type and Surgical Volume on 10-Year Survival in Patients Undergoing Hepatic Resection for Hepatocellular Carcinoma. *J Am Coll Surg* 2017, 224 (3), 362-372.
34. Sheetz, K. H.; Dimick, J. B.; Ghaferi, A. A., Impact of Hospital Characteristics on Failure to Rescue Following Major Surgery. *Ann Surg* 2016, 263 (4), 692-7.
35. Ghaferi, A. A.; Birkmeyer, J. D.; Dimick, J. B., Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg* 2009, 250 (6), 1029-34.
36. McMurry, T. L.; Shah, P. M.; Samson, P.; Robinson, C. G.; Kozower, B. D., Treatment of stage I non-small cell lung cancer: What's trending? *J Thorac Cardiovasc Surg* 2017, 154 (3), 1080-1087.
37. Bilimoria, K. Y.; Bentrem, D. J.; Ko, C. Y.; Stewart, A. K.; Winchester, D. P.; Talamonti, M. S., National failure to operate on early stage pancreatic cancer. *Ann Surg* 2007, 246 (2), 173-80.
38. Luryi, A. L.; Chen, M. M.; Mehra, S.; Roman, S. A.; Sosa, J. A.; Judson, B. L., Hospital readmission and 30-day mortality after surgery for oral cavity cancer: Analysis of 21,681 cases. *Head Neck* 2016, 38 *Suppl 1*, E221-6.
39. Wegner, R. E.; Verma, V.; Hasan, S.; Schiffman, S.; Thakkar, S.; Horne, Z. D.; Kulkarni, A.; Williams, H. K.; Monga, D.; Finley, G.; Kirichenko, A. V., Incidence and risk factors for post-operative mortality, hospitalization, and readmission rates following pancreatic cancer resection. *J Gastrointest Oncol* 2019, 10 (6), 1080-1093.
40. Barber, E. L.; Rossi, E. C.; Gehrig, P. A., Surgical readmission and survival in women with ovarian cancer: Are short-term quality metrics incentivizing decreased long-term survival? *Gynecol Oncol* 2017, 147 (3), 607-611.
41. Kamarajah, S. K.; Burns, W. R.; Frankel, T. L.; Cho, C. S.; Nathan, H., Validation of the American Joint Commission on Cancer (AJCC) 8th Edition Staging System for Patients with Pancreatic Adenocarcinoma: A Surveillance, Epidemiology and End Results (SEER) Analysis. *Ann Surg Oncol* 2017, 24 (7), 2023-2030.
42. Kamarajah, S. K.; Frankel, T. L.; Sonnenday, C.; Cho, C. S.; Nathan, H., Critical evaluation of the American Joint Commission on Cancer (AJCC) 8th edition staging system for patients with Hepatocellular Carcinoma (HCC): A Surveillance, Epidemiology, End Results (SEER) analysis. *J Surg Oncol* 2018, 117 (4), 644-650.
43. Kamarajah, S. K., Evaluation of the AJCC 8th Edition Staging System for Pathologically Versus Clinically Staged Intrahepatic Cholangiocarcinoma (iCCA): a Time to Revisit a Dogma? A Surveillance, Epidemiology, and End Results (SEER) Analysis. *J Gastrointest Cancer* 2018.
44. Fong, Y.; Wagman, L.; Gonen, M.; Crawford, J.; Reed, W.; Swanson, R.; Pan, C.; Ritchey, J.; Stewart, A.; Choti, M., Evidence-based gallbladder cancer staging: changing cancer staging by analysis of data from the National Cancer Database. *Ann Surg* 2006, 243 (6), 767-71; discussion 771-4.
45. Weksler, B.; Kennedy, K. F.; Sullivan, J. L., Using the National Cancer Database to create a scoring system that identifies patients with early-stage esophageal cancer at risk for nodal metastases. *J Thorac Cardiovasc Surg* 2017, 154 (5), 1787-1793.
46. Nathan, H.; Hyder, O.; Mayo, S. C.; Hirose, K.; Wolfgang, C. L.; Choti, M. A.; Pawlik, T. M., Surgical therapy for early hepatocellular carcinoma in the modern era: a 10-year SEER-medicare analysis. *Ann Surg* 2013, 258 (6), 1022-7.
47. Paniccia, A.; Hosokawa, P.; Henderson, W.; Schulick, R. D.; Edil, B. H.; McCarter, M. D.; Gajdos, C., Characteristics of 10-Year Survivors of Pancreatic Ductal Adenocarcinoma. *JAMA Surg* 2015, 150 (8), 701-10.
48. Little, R. J. A.; Rubin, D. B., *Statistical analysis with missing data*. 2nd ed. Hoboken NJ ed.; Wiley: 2002.
49. Schafer, J. L., Multiple imputation: a primer. *Stat Methods Med Res* 1999, 8 (1), 3-15.
50. Nathan, H.; Pawlik, T. M.; Wolfgang, C. L.; Choti, M. A.; Cameron, J. L.; Schulick, R. D., Trends in survival after surgery for cholangiocarcinoma: a 30-year population-based SEER database analysis. *J Gastrointest Surg* 2007, 11 (11), 1488-96; discussion 1496-7.
51. Nathan, H.; Pawlik, T. M., Staging of intrahepatic cholangiocarcinoma. *Curr Opin Gastroenterol* 2010, 26 (3), 269-73.
52. Nathan, H.; Pawlik, T. M., Limitations of claims and registry data in surgical oncology research. *Ann Surg Oncol* 2008, 15 (2), 415-23.
53. de Jong, M. C.; Nathan, H.; Sotiropoulos, G. C.; Paul, A.; Alexandrescu, S.; Marques, H.; Pulitano, C.; Barroso, E.; Clary, B. M.; Aldrighetti, L.; Ferrone, C. R.; Zhu, A. X.; Bauer, T. W.; Walters, D. M.; Gamblin, T. C.; Nguyen, K. T.; Turley, R.; Popescu, I.; Hubert, C.; Meyer, S.; Schulick, R. D.; Choti, M. A.; Gigot, J. F.; Mentha, G.; Pawlik, T. M., Intrahepatic cholangiocarcinoma: an international multi-institutional analysis of prognostic factors and lymph node assessment. *J Clin Oncol* 2011, 29 (23), 3140-5.
54. Mantripragada, K. C.; Hamid, F.; Shafqat, H.; Olszewski, A. J., Adjuvant Therapy for Resected Gallbladder Cancer: Analysis of the National Cancer Data Base. *J Natl Cancer Inst* 2017, 109 (2).
55. Mitin, T.; Enestvedt, C. K.; Jemal, A.; Sineshaw, H. M., Limited Use of Adjuvant Therapy in Patients With Resected Gallbladder Cancer Despite a Strong Association With Survival. *J Natl Cancer Inst* 2017, 109 (7).
56. Delgado-Rodriguez, M.; Llorca, J., Bias. *J Epidemiol Community Health* 2004, 58 (8), 635-41.
57. Rubin, D. B., Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997, 127 (8 Pt 2), 757-63.
58. Braitman, L. E.; Rosenbaum, P. R., Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002, 137 (8), 693-5.
59. Grose, E.; Wilson, S.; Barkun, J.; Bertens, K.; Martel, G.; Balaa, F.; Abou Khalil, J., Use of Propensity Score Methodology in

Contemporary High-Impact Surgical Literature. *J Am Coll Surg* 2020, 230 (1), 101-112 e2.

60. Kamarajah, S. K.; Bednar, F.; Cho, C. S.; Nathan, H., Survival benefit with adjuvant radiotherapy after resection of distal cholangiocarcinoma: A propensity-matched National Cancer Database analysis. *Cancer* 2020.
61. Mathes, T.; Buehn, S.; Prengel, P.; Pieper, D., Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. *J Clin Epidemiol* 2018, 93, 120-127.