HHS Public Access

Author manuscript

Pediatr Blood Cancer. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Pediatr Blood Cancer. 2015 September ; 62(9): 1495–1500. doi:10.1002/pbc.25506.

Use of Patient Registries and Administrative Datasets for the Study of Pediatric Cancer

Henry E. Rice, MD[1,2,*], **Brian R. Englum, MD**[1], **Brian C. Gulack, MD**[1], **Obinna O. Adibe, MD**[1,2], **Elizabeth T. Tracy, MD**[1,2], **Susan G. Kreissman, MD**[2], and **Jonathan C. Routh, MD**[1,2]

[1]Department of Surgery, Duke University Medical Center, Durham, North Carolina

[2]Department of Pediatrics, Duke University Medical Center, Durham, North Carolina

Abstract

Analysis of data from large administrative databases and patient registries is increasingly being used to study childhood cancer care, although the value of these data sources remains unclear to many clinicians. Interpretation of large databases requires a thorough understanding of how the dataset was designed, how data were collected, and how to assess data quality. This review will detail the role of administrative databases and registry databases for the study of childhood cancer, tools to maximize information from these datasets, and recommendations to improve the use of these databases for the study of pediatric oncology.

Keywords

administrative dataset; patient registry; pediatric cancer; secondary analysis

INTRODUCTION

Data from national cooperative study groups, such as the Children's Oncology Group (COG) in the U.S. or the International Society for Pediatric Oncology (SIOP) in Europe, have helped define much of the evidence to guide pediatric cancer treatment [1,2]. Despite the value of cooperative study groups, there are several limitations to use of their data for research, including a high administrative burden, lack of detailed cost data, and difficulty performing comparative effectiveness research beyond the primary clinical trial question [3]. Although COG operates the Children's Cancer Research Network registry [4], data from cooperative groups have limited application to the study of population-based cancer incidence, as they tend to focus on patients from participating centers following enrollment in individual trials [5–8].

The use of administrative databases and patient cancer and surgical registries provide alternative approaches to the study of childhood cancer. Administrative databases are often referred to as "secondary data," because they are collected primarily for non-research

*Correspondence to: Henry E. Rice, Division of Pediatric Surgery, Box 3815, Duke University Medical Center, Durham, NC 27710. rice0017@mc.duke.edu.

Conflict of interest: Nothing to declare.

purposes. Although these databases are not focused on the study of cancer, they do include important outcomes such as length of stay, charges and/or costs, and some adverse events [3]. In general, administrative databases tend to be good at providing broad overviews of healthcare practices, such as variation in practice patterns, outcome–volume relationships, or care trends over time [9].

Patient registries focus on a particular disease or intervention, often include extended follow-up [10]. Although registries can be used to study a wide range of outcomes (including ideas that were not conceived during initial registry design), they may suffer from a lack of standardized data elements, difficulty assessing trends in care, or lack of control groups. In general, registry studies tend to be most successful when analyzing a specific research question for the population for which they were initially designed.

With the increasing use of administrative databases and patient registries to study pediatric cancer care, there is some skepticism about the value of these sources, in particular about overreaching conclusions or inappropriate analytic methods [9,11]. These are legitimate concerns, as interpretation of these datasets requires a thorough understanding of database design, data quality, and how outcomes of interest were recorded. This review will describe commonly used administrative databases and registries, tools to manage information from these datasets, and a list of recommendations to improve the use of these databases for pediatric cancer research. This report is not an exhaustive or systematic review of all available datasets, rather a “how-to” guide to assist the pediatric clinician with understanding of available tools. As well, although beyond the scope of this review, the use of “big data” to understand the biology of cancer such as through the National Cancer Genome Atlas is rapidly becoming part of personalized cancer care for children.

ADMINISTRATIVE DATABASES

Healthcare Cost and Utilization Project

The Healthcare Cost and Utilization Project (HCUP) is a family of databases sponsored by the Agency for Healthcare Research and Quality (Table I) representing the largest collection of hospital data in the U.S. [12]. These databases are particularly useful for the study of treatment complications, in-hospital mortality, length of stay, practice patterns, health care disparities, and costs.

State Inpatient, Emergency, and Ambulatory Databases

The State Inpatient Databases (SID) and the State Ambulatory Surgery Databases (SASD) are complementary HCUP datasets which allow the comparison of outcomes from inpatient or outpatient settings, respectively [13,14]. Currently, 47 states provide discharge data to SID, capturing approximately over 95% of all U.S. hospital discharges; 33 states contribute out-patient data to SASD. SID and SASD data are re-coded by HCUP into a consistent format that allows intra- and inter-state comparisons. Importantly, most states abstract SID and SASD data as per-encounter events, not per-patient events; however, some states allow researchers to track patients longitudinally. This allows researchers to assess important oncologic outcomes such as unplanned readmissions or emergency department visits.

National (Nationwide) Inpatient Sample

The National (Nationwide) Inpatient Sample (NIS) is a derivative dataset developed by HCUP using SID data, and is the largest publicly available all-payer inpatient health care database in the U.S. It contains data from more than 7 million pediatric and adult admissions annually from over 1,000 hospitals, representing ~20% of discharges in participating states [15]. NIS is ideally suited for investigating trends in care utilization or healthcare costs, variation in practice, or diffusion of new technologies. Advantages of NIS include its population-based design, consistency over time, large size, and detailed cost data.

However, NIS has several disadvantages for the study of pediatric cancer. As with most administrative datasets, there is limited collection of cancer-specific variables, such as tumor staging, grade, and pathology. Unlike state databases, there is no longitudinal follow-up of individual patients and no outpatient data. As well, although NIS does provide some data to estimate disease incidence and cancer-specific-mortality, it is overall limited for these uses. Finally, NIS (like most administrative datasets) tends to both underestimate mortality and overestimate morbidity, as these outcomes are only measured at discharge. For example, Ambekar et al. [16] have shown in children with spinal meningioma who were disabled before surgery were more likely to get discharged to a facility regardless of the effect of surgery, thus overestimating the morbidity of surgery.

Kids Inpatient Database

The Kids Inpatient Database (KID) is a derivative dataset compiled by HCUP that is an enriched sample of inpatient data for children under 21 years old. KID is based on a triennial systematic sample of discharges from over 5,000 hospitals across the U.S., including short-term, non-Federal, general and specialty hospitals [17]. KID is often used for development of national hospitalization requirements, resource utilization, or factors associated with specific diagnoses or procedures. For example, one ideal use of KID for the study of pediatric cancer was by Chu et al. [18], who identified factors associated with radical nephrectomy or nephron-sparing surgery for children with renal cancer.

Similar sampling strata are used by NIS and KID, with addition of a stratifier in KID that identifies pediatric hospitals. The addition of this modifier facilitates studies of variations in clinical outcomes and resource utilization between pediatric and non-pediatric hospitals [19]. However, as with NIS, KID uses hospital admissions as the primary unit of data; thus repeat admissions cannot be accounted for. Similarly, diagnoses and procedures are restricted to billing codes.

OTHER ADMINISTRATIVE DATABASES

Pediatric Health Information System

The Pediatric Health Information System (PHIS) is a privately administered database that contains data from inpatient admissions, ambulatory and short-stay encounters, and emergency rooms from over 40 children’s hospitals in the U.S. PHIS provides greater granularity than the HCUP databases. Particularly for surgical care, PHIS has been used to study practice variability and health care utilization [20]. PHIS can be used to follow a child

over encounters using unique identifiers, although not all outpatient encounters are captured. Because of this longitudinal capacity, PHIS is a popular database for research in pediatric care, although with all private datasets additional costs are required.

Similar to most administrative databases, PHIS is limited by its use of International Classification of Diseases, 9th edition (ICD-9) diagnosis codes and risk of coding errors, which have been shown to range from 2% to 4% in a nested chart review of children with renal cancer [20]. To overcome this challenge, Desai et al. developed an algorithm (using ICD-9 codes, exclusion criteria, and manual review of chemotherapy billing data) to assemble a high-risk neuroblastoma cohort at a single institution [21]. Similarly, Kavcic et al. [22] confirmed an algorithm to identify children using PHIS with acute myeloid leukemia (AML) based on ICD-9 diagnosis and manual review of chemotherapy.

As researchers and clinicians prepare for the transition from ICD-9 to ICD-10 over the next several years, the impact on use of PHIS and other datasets is unclear. This transition will increase the specificity but also complexity of billing code-derived data, potentially increasing the risk of coding errors. For instance, while ICD-9 contains 14,000+ diagnostic codes, ICD-10 will contain 90,000+ codes with built in flexibility for adding new codes. For experienced researchers, this transition will hopefully allow for increased sophistication in analysis of specific questions using administrative datasets.

Insurer-Specific Datasets

Many large insurers collect claims data to help manage their programs. These datasets (e.g., MarketScan, i3 Innovus, Harvard Pilgrim, Kaiser, etc.) provide high-quality patient-level data from commercial, Medicare supplemental, and Medicaid populations. These datasets provide “real-world” treatment patterns, and are most useful for cost analyses, including longitudinal analysis. However, these data sets are proprietary, and access can be prohibitively expensive for many researchers. For children’s cancer care, additional disadvantages are that small populations within a single insurer may not be nationally representative; patients may pay out of pocket and not file a claim; and patients may switch insurance companies, limiting longitudinal follow-up [9].

Medicaid Analytic Extract (MAX)

Each state’s Medicaid or Children’s Health Insurance Program (CHIP) data are collected nationally by the Medicaid and CHIP Statistical Information System (MSIS). CMS operates the Medicaid Analytic Extract (MAX) dataset, which accounts for people enrolled after retroactive corrections have been applied and after state-specific data elements have been transformed into a consistent format. The claims in MAX can identify a broad picture of clinical services rendered and the cost of services after adjustments. These data are most helpful to summarize financial data, service utilization, and care expenditures for a particular diagnosis or procedure.

CANCER AND SURGICAL REGISTRIES

Surveillance, Epidemiology, and End Results Database

The Surveillance, Epidemiology and End Results (SEER) database is a population-based cancer registry that compiles data from hospital and state cancer registries. SEER covers over 25% of the U.S. population, and captures demographic and cancer-specific variables, including disease incidence, histology, staging, radiation dosing, surgical extent, and cancer-specific mortality [23]. One particular role for this dataset is the identification of a population-based cohort of children with cancer, which allows for identification of factors which affect outcomes. For example, Potonski et al. showed that SEER can be used to identify a cohort of adolescent and young adult children with various cancers to identify factors associated with appropriate treatment [24].

SEER, along with the National Cancer Data Base and the National Program of Cancer Registries (detailed below), operates through similar coding procedures as described in the Facility Oncology Registry Data Standards (FORDS) manual. However, as these groups' data is based on American Joint Commission on Cancer (AJCC) staging, their value is limited for pediatric cancers that use other staging systems, such as neuroblastoma and rhabdomyosarcoma. In addition, chemotherapy, comorbidity data, and other clinical factors beyond first-line therapy have limited availability.

National Cancer Data Base

The National Cancer Data Base (NCDB) is a national cancer registry administered by the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society. With a catchment from over 1,500 centers who participate in the CoC, NCDB has data from over 30 million patients, with a focus on adults with site-specific cancers, such as those of the breast, colon, and lung. For children, NCDB data has been increasingly used to identify patters of disease incidence and survival for pediatric cancers [25].

There are several limitations to the NCDB for the study of pediatric cancer. First, it is based on convenience sampling from participating centers, and it is not ideal for examining disease trends over time. Second, similar to SEER, many pediatric disease specific variables are not collected (e.g., N-MYC amplification for neuroblastoma). Finally, many freestanding children's hospitals are not members of the CoC, and data from these children are not included. However, these limitations are recognized by the leadership of the NCDB and SEER, with ongoing revisions in progress to incorporate contemporary staging, treatment, and histology variables for pediatric cancer.

National Program of Cancer Registries

The National Program of Cancer Registries (NPCR) is a program operated by the Centers for Disease Control (CDC) to support state cancer registries with data management. In line with the public health goals of the CDC, several NPCR reports are compiled annually of national cancer surveillance data, including the United States Cancer Statistics (USCS): Incidence and Mortality report of official federal government statistics for new cancer cases

and deaths, the Interactive Cancer Atlas (http://apps.nccd.cdc.gov/DCPC_INCA/DCPC_INCA.aspx), the CDC WONDER online query system for age-adjusted and crude cancer rates (http://wonder.cdc.gov/cancer.html), and the U.S. County Cancer Incidence Dataset of aggregate cancer incidence rates for select U.S. counties.

For the study of children's cancer, similarities in coding between SEER and the NPCR allow for complementary study of large populations, particularly for epidemiologic analysis. For example, Siegel et al. analyzed data from the NPDS and SEER to capture statewide registries representing 94.2% of the US population to identify the rates of various cancers in children [26]. Limitations of the NPCR, similar to SEER, include that treatment type is limited to first course and even then, some chemotherapy data is missing, limited information on co-morbidities, and outcome data is generally limited to survival, making other important outcomes such a tumor recurrence difficult to track [27]. Finally, duplication or cross-over of subjects between cancer registries is difficult to identify, limited the ability to merge data.

Children's Cancer Research Network

The Childhood Cancer Research Network (CCRN) registry was established by COG in 2008 to increase registry data at COG centers. The CCRN enrolls children with cancer who are treated at a COG institution in the U.S. and Canada through an additional informed consent process [4]. Although the research experience with the CCRN remains limited to date, it does offer a centralized resource for many areas of cancer research.

Similar to the NCDB, limitations of CCRN include bias from convenience sampling, as children are only enrolled after diagnosis at a COG institution, as well as gaps in ascertainment of subgroups based on age, diagnosis, sex, race/ethnicity, or geographic region [4]. Despite these concerns, Musselman et al. demonstrated a high rate of catchment in the CCRN by comparison with expected number of cases from SEER and U.S. Census data [28]. Overall, 42% of predicted children with cancer at participating institutions were registered in the CCRN, with some malignancies better represented (leukemia, 59%; renal tumors, 67%) than others (retinoblastoma, 34%).

National Surgical Quality Improvement Program-Pediatric

The National Surgical Quality Improvement Program-Pediatric (NSQIP-Ped) is a pediatric surgical registry operated by the American College of Surgeons. This program collects standardized perioperative (30-day) outcome data from centers across the U.S. The program provides risk-adjusted outcomes and comparison to other institutions back to participating centers to promote quality improvement initiatives, and de-identified datasets are available to participating centers to promote clinical research. However, NSQIP-Ped is limited for the study of pediatric cancer by its lack of staging information, cancer-specific outcomes (e.g., surgical margins, nodal harvest, disease recurrence), and longitudinal follow-up.

TOOLS TO ENHANCE LARGE DATASET ANALYSIS

Merging Datasets

One approach that is gaining interest among pediatric researchers is the use of merged datasets. This approach leverages the strengths of multiple data sources to overcome the limitations of each. For example, registries often contain detailed disease-specific information but lack data concerning resource utilization. Administrative databases tend to excel at capturing data on resource utilization, but often do not contain disease-specific information. There are several limitations to the merging of datasets. For example, although some datasets collect patient identifiers, these are not routinely distributed to researchers or made publicly available because of privacy concerns. Other limitations include the high administrative burden and complexity of algorithms to merge cohorts.

Several pediatric groups have recently described innovative methods to link datasets. Aplenc et al. detailed an approach for merging COG clinical data with PHIS data for children with AML [3]. Pasquali et al. [29] described linkage of data for children with congenital heart disease from The Society of Thoracic Surgeons Congenital Heart Surgery Database with PHIS using indirect identifiers and probabilistic matching. Deans et al. [30] validated a similar algorithm using indirect identifiers to link data from NSQIP-Peds and PHIS to investigate healthcare utilization during the first post-operative year.

Clinical Classification Software (CCS)

As administrative datasets generally rely on discharge diagnosis and billing codes, these datasets pose challenges for studying cancer in which multiple diagnoses and codes are often pertinent for a single care encounter. To address this limitation, HCUP developed Clinical Classification Software (CCS), a tool which groups thousands of ICD-9 codes into 260 diagnostic groups and 231 procedure groups [31,32]. For the study of childhood cancer, CCS provides an organized categorization scheme that collapses multiple ICD-9-CM codes into a smaller number of clinically meaningful categories.

Russell et al. [33] described use of CCS to classify cancer-related admissions from the KID dataset into categories that reflect real-world clinical practice. This group used a multistep process for stratifying cancer-related admissions into four categories: chemotherapy-related, procedure-related, infection-related, or toxicity-related. Similarly, Mueller et al. described the use of CCS to describe the reasons for emergency department visits among pediatric cancer patients, including risk factors for admission to the hospital [34].

Comorbidity Tools

In contrast to clinical trials, when using data from administrative databases, the effect of preexisting conditions (comorbidities) may be difficult to control, leading to difficulty with non-random treatment selection. In general, comorbidities have been handled analytically in these settings by: (1) stratifying patients into groups-those with a comorbidity and those without; (2) using separate binary indicators for discrete conditions; or (3) summarizing comorbidity information into a score that provides a single parameter for measuring multiple comorbidities [35].

Several methods are available to assist with managing comorbidities in large administrative datasets. Comorbidity Software is a tool developed by HCUP to help address these issues, as it assigns variables for comorbidities in discharge records using ICD-9-CM codes [35]. This software has been shown to increase the identification of comorbidities and separate them from the primary reason for hospitalization [36].

Analytic methods to account for non-random treatment selection include propensity score analysis, inverse probability weighting and logistic regression, among others. In particular, propensity score analysis is increasingly being applied to the study of pediatric cancer. Seif et al. [37] used propensity scoring to estimate the secondary AML risk in children receiving dexrazoxane after anthracycline exposure using the PHIS. Wilson et al. [38] used propensity score analysis to determine the attributable cost and length of stay of central line-associated bloodstream infections in children with cancer while controlling for covariates.

LIMITATIONS OF REGISTRIES AND ADMINISTRATIVE DATABASES

Cancer outcomes can only be understood using large datasets if there is accurate coding of diagnoses and procedures [39]. The most typical types of error are: (1) overlooking of diagnoses; (2) incorrect or skipped induction; (3) indexing errors; (4) violation of ICD rules and external regulations [39]. These errors generally result from mistakes in the primary documentation, insufficient knowledge of the encoders and registrars (different steps of the coding process require different kind of knowledge), and internal inconsistency among multiple codes. Coding validation is performed in several administrative datasets and patient registries, but is not universally practiced.

One particular limitation of administrative datasets relevant for the study of pediatric cancer is difficulty with tracking of adverse events. As administrative datasets tend to focus on inpatient data, they may underestimate the long-term risk of adverse events, as many cancer-related adverse events occur after discharge (e.g., surgical site infections or pulmonary emboli) [40]. Linkage of clinical datasets to administrative datasets may offer the greatest potential to improve the accuracy of adverse events tracking in pediatric cancer.

SUMMARY AND RECOMMENDATIONS

In summary, the use of administrative databases and patient registries offers tremendous opportunities to enhance our understanding of pediatric cancer care through the systematic study of large numbers of patients. Used correctly, these resources can help us better understand cancer treatment outcomes, quality of care, resource utilization, and clinical management.

For the novice in use of these datasets, several steps are helpful to facilitate the study of pediatric cancer. Researchers using registries and administrative databases must remember that the data they are using was not collected in order to answer their specific research question. The first step in designing a study is to determine whether a specific question can be adequately answered by a given database. A good tactic to address this issue is to collaborate with an experienced team who has previously used that specific database. Clinicians should work closely with biostatisticians, informaticians, and programmers to

confirm that their analyses are showing what they think they are showing. Secondly, prior to collection of data, the research team should ensure that the underlying research questions are important, that the dataset is appropriate to address these issues, and that the statistical methods are valid to answer the particular research questions [9].

To improve the use of administrative datasets and patient registries for the study of pediatric cancer, we offer the following recommendations:

1. One advantage of administrative databases is that they capture economic data well. However, use of these datasets to define the costs of pediatric cancer care remains largely ignored. Improved study of key economic information, such as direct and indirect costs of care, resource utilization, as well as opportunity costs of delayed or absent care is essential to guide policy development.
2. Specific to pediatric cancer, patient and tumor-specific variables (e.g., N-MYC status for neuroblastoma) should be better integrated into national cancer registries. Pediatric cancer registries would greatly benefit from inclusion of contemporary disease-specific staging, pathology, and outcomes, as is currently being developed for the NCDB and SEER through the FORDS revision project. As well, large pediatric hospitals who are not in the Commission on Cancer (and therefore do not participate in the NCDB) should more actively participate in these efforts.
3. Collaboration between the leadership of cooperative study groups and administrative databases to work toward a single high-quality national cancer database system may best facilitate the study of pediatric cancer for outcomes research [27]. Federal and organizational support will be required for such an comprehensive effort, as private insurers and other groups do not have the resources or motivation to work toward this overarching goal.
4. To minimize coding errors, validation studies should be performed whenever possible to verify that data are coded accurately and that observed trends are real. Alternatively, it can be useful to ask the same research question of different databases in order to verify that the results are consistent, as has been tested in pediatric renal cancer [20]. Put differently, researchers should make sure that they are detecting signal, not noise.
5. Journal editors and reviewers have an increasing responsibility to ensure the validity of analyses using large datasets. Development of guidelines for secondary data analysis should address standards for data quality, content, analytic methods, standardization, and interpretation. Similar guidelines have become an integral part of the clinical research process, such as advocated by the Consolidated Standards of Reporting Trials (CONSORT) statement (www.consort-statement.org) [41].

REFERENCES

1. Nass, SJ.; Moses, HL.; Mendelsohn, J. A National Cancer Clinical Trials System for the 21st Century; Reinvigorating the NCI Cooperative Group Program. Washington, DC: Institute of Medicine; 2010.

2. Mott MG, Mann JR, Stiller CA. The United Kingdom Children's Cancer Study Group—The first 20 years of growth and development. Eur J Cancer. 1997; 33:1448–1452. [PubMed: 9337688]
3. Aplenc R, Fisher BT, Huang YS, Li Y, Alonzo TA, Gerbing RB, Hall M, Bertoch D, Keren R, Seif AE, Sung L, Adamson PC, Gamis A. Merging of the National Cancer Institute-funded cooperative oncology group data with an administrative data source to develop a more effective platform for clinical trial analysis and comparative effectiveness research: A report from the Children's Oncology Group. Pharmacoepidemiol Drug Saf. 2012; 21:37–43. [PubMed: 22552978]
4. Steele JR, Wellemeyer AS, Hansen MJ, Reaman GH, Ross JA. Childhood Cancer Research Network: A North American pediatric cancer registry. Cancer Epidemiol Biomarkers Prev. 2006; 15:1241–1242. [PubMed: 16835317]
5. Fisher B, Harris T, Torp K, Seif A, Shah A, Huang Y, Bailey L, Kersun L, Reilly A, Rheingold S, Walker D, Li Y, Aplenc R. Establishment of an 11-year cohort of 8733 pediatric patients hospitalized at United States free-standing children's hospitals with de novo acute lymphoblastic leukemia from health care administrative data. Med Care. 2014; 52:e1–e6. [PubMed: 22410405]
6. Kavcic M, Fisher B, Li Y, Seif A, Torp K, Walker D, Huang Y, Lee G, Tasian S, Vujkovic M, Bagatell R, Aplenc R. Induction mortality and resource utilization in children treated for acute myeloid leukemia at free-standing pediatric hospitals in the United States. Cancer. 2013; 119:1916–1923. [PubMed: 23436301]
7. Shaw P, Ritchey A. Different rates of clinical trial enrollment between adolescents and young adults aged 15 to 22 years and children under 15 years old with cancer at a children's hospital. J Pediatr Hematol Oncol. 2007; 29:811–814. [PubMed: 18090927]
8. Koschmann C, Thomson B, Hawkins D. No evidence of a trial effect in newly diagnosed pediatric acute lymphoblastic leukemia. Arch Pediatr Adolesc Med. 2010; 164:214–217. [PubMed: 20194252]
9. Schlomer BJ, Copp HL. Secondary data analysis of large data sets in urology: Successes and errors to avoid. J Urol. 2014; 191:587–596. [PubMed: 24140846]
10. Gliklich, R.; Dreyer, N., editors. Registries for evaluating patient outcomes: A user's guide. 2nd edition. Rockville, MD: Agency for Healthcare Research and Quality; 2010. (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc., d/b/a Outcome] under Contract No. HHSA29020050035I T03). AHRQ Publication No. 10-EHC049
11. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. J Clin Epidemiol. 2007; 60:734–741. [PubMed: 17573990]
12. Agency for Healthcare Research and Quality (AHRQ). [cited February 1, 2008] Overview of HCUP (Healthcare Cost and Utilization Project). Available from: http://www.hcup-us.ahrq.gov/overview.jsp
13. Agency for Healthcare Research and Quality (AHRQ). [cited February 20, 2008] Overview of the State Inpatient Databases (SID). Available from: http://www.hcup-us.ahrq.gov/sidoverview.jsp
14. Agency for Healthcare Research and Quality (AHRQ). [cited February 20, 2008] Overview of the State Ambulatory Surgery Databases (SASD). Available from: http://www.hcup-us.ahrq.gov/sasdoverview.jsp
15. Agency for Healthcare Research and Quality (AHRQ). [cited February 20, 2008] Overview of the Nationwide Inpatient Sample (NIS). Available from: http://www.hcup_us.ahrq.gov/nisoverview.jsp
16. Ambekar S, Sharma M, Kukreja S, Nanda A. Complications and outcomes of surgery for spinal meningioma: A Nationwide Inpatient Sample analysis from 2003 to 2010. Clin Neurol Neurosurg. 2014; 118:65–68. [PubMed: 24529232]
17. Agency for Healthcare Research and Quality (AHRQ). [cited February 20, 2008] Overview of the Kids' Inpatient Database (KID). Available from: http://www.hcup-us.ahrq.gov/kidoverview.jsp
18. Chu DI, Lloyd JC, Balsara ZR, Wiener JS, Ross SS, Routh JC. Variation in use of nephron-sparing surgery among children with renal tumors. J Pediatr Urol. 2014; 10:724–729. [PubMed: 24517904]
19. Odetola FO, Gebremariam A, Freed GL. Patient and hospital correlates of clinical outcomes and resource utilization in severe pediatric sepsis. Pediatrics. 2007; 119:487–494. [PubMed: 17332201]

20. Routh JC, Graham DA, Estrada CR, Nelson CP. Contemporary use of nephron-sparing surgery for children with malignant renal tumors at freestanding children's hospitals. Urology. 2011; 78:422–426. [PubMed: 21689846]
21. Desai AV, Kavcic M, Huang Y-S, Herbst N, Fisher BT, Seif AE, Li Y, Hennessy S, Aplenc R, Bagatell R. Establishing a high-risk neuroblastoma cohort using the pediatric health information system database. Pediatr Blood Cancer. 2014; 61:1129–1131. [PubMed: 24616331]
22. Kavcic M, Fisher BT, Torp K, Li Y, Huang Y-S, Seif AE, Vujkovic M, Aplenc R. Assembly of a cohort of children treated for acute myeloid leukemia at free-standing children's hospitals in the United States using an administrative database. Pediatr Blood Cancer. 2013; 60:508–511. [PubMed: 23192853]
23. Ries, LAG.; Melbert, D.; Krapcho, M.; Stinchcomb, DG.; Howlader, N.; Horner, MJ.; Mariotto, A.; Miller, BA.; Feuer, EJ.; Altekruse, SF.; Lewis, DR.; Clegg, L.; Eisner, MP.; Reichman, M.; Edwards, BK. SEER cancer statistics review, 1975–2005. Bethesda, MD: National Institutes of Health, National Cancer Institute; 2008.
24. Potosky AL, Harlan LC, Albritton K, Cress RD, Friedman DL, Hamilton AS, Kato I, Keegan THM, Keel G, Schwartz SM, Seibel NL, Shnorhavorian M, West MM, Wu X-C. Group tAHSC. Use of appropriate initial treatment among adolescents and young adults with cancer. J Natl Cancer Inst. 2014; 106
25. Grovas A, Fremgen A, Rauck A, Ruymann FB, Hutchinson CL, Winchester DP, Menck HR. The National Cancer Data Base report on patterns of childhood cancers in the United States. Cancer. 1997; 80:2321–2332. [PubMed: 9404710]
26. Siegel DA, King J, Tai E, Buchanan N, Ajani UA, Li J. Cancer incidence rates and trends among children and adolescents in the United States, 2001–2009. Pediatrics. 2014; 134:e945–e955. [PubMed: 25201796]
27. Gotay, CC.; Lipscomb, J. Data for cancer outcomes research: Identifying and strengthening the empirical base. In: Lipscomb, J.; Gotay, CC.; Snyder, C., editors. Outcomes assessment in cancer: Measure, methods, and applications. Cambridge: Cambridge University Press; 2005. p. 533-549.
28. Musselman JRB, Spector LG, Krailo MD, Reaman GH, Linabery AM, Poynter JN, Stork SK, Adamson PC, Ross JA. The Children's Oncology Group Childhood Cancer Research Network (CCRN): Case catchment in the United States. Cancer. 2014; 120:3007–3015. [PubMed: 24889136]
29. Pasquali SK, Jacobs JP, Shook GJ, O'Brien SM, Hall M, Jacobs ML, Welke KF, Gaynor JW, Peterson ED, Shah SS, Li JS. Linking clinical registry data with administrative data using indirect identifiers: Implementation and validation in the congenital heart surgery population. Am Heart J. 2010; 160:1099–1104. [PubMed: 21146664]
30. Deans KJ, Cooper JN, Rangel SJ, Raval MV, Minneci PC, Moss RL. Enhancing NSQIP-Pediatric through integration with the Pediatric Health Information System. J Pediatr Surg. 2014; 49:207–212. [PubMed: 24439611]
31. Woodworth G, Baird C, Garces-Ambrossi G, Tonascia J, Tamargo R. Inaccuracy of the administrative database: Comparative analysis of two databases for the diagnosis and treatment of intracranial aneurysms. Neurosurgery. 2009; 65:251–256. [PubMed: 19625902]
32. Elixhauser, A.; Steiner, C.; Palmer, L. Clinical classifications software (CCS). Rockville, MD: Agency for Healthcare Research and Quality; 2004.
33. Russell H, Okcu M, Kamdar K, Shah M, Kim E, Swint J, Chan W, Du X, Franzini L, Ho V. Algorithm for analysis of administrative pediatric cancer hospitalization data according to indication for admission. BMC Med Inform Decis Mak. 2014; 14:88. [PubMed: 25274165]
34. Mueller EL, Sabbatini A, Gebremariam A, Mody R, Sung L, Macy ML. Why pediatric patients with cancer visit the emergency department: United States, 2006–2010. Pediatr Blood Cancer. 2015; 62:490–495. [PubMed: 25345994]
35. Elixhauser A, Steiner C, Harris DR, Coffey R. Comorbidity measures for use with administrative data. Med Care. 1998; 36:8–27. [PubMed: 9431328]
36. Reeve BB, Smith AW, Arora NK, Hays RD. Reducing bias in cancer research: Application of propensity score matching. Health Care Financ Rev. 2008; 29:69–80. [PubMed: 18773615]

37. Seif AE, Walker DM, Li Y, Huang Y-SV, Kavcic M, Torp K, Bagatell R, Fisher BT, Aplenc R. Dexrazoxane exposure and risk of secondary acute myeloid leukemia in pediatric oncology patients. Pediatr Blood Cancer. 2014; 62:704–709. [PubMed: 24668949]

38. Wilson MZ, Rafferty C, Deeter D, Comito MA, Hollenbeak CS. Attributable costs of central line–associated bloodstream infections in a pediatric hematology/oncology population. Am J Infect Control. 2014; 42:1157–1160. [PubMed: 25444262]

39. Surján G. Questions on validity of International Classification of Diseases-coded diagnoses. Int J Med Inform. 1999; 54:77–95. [PubMed: 10219948]

40. Bilimoria KY, Cohen ME, Ingraham AM, Bentrem DJ, Richards K, Hall BL, Ko CY. Effect of postdischarge morbidity and mortality on comparisons of hospital surgical quality. Ann Surg. 2010; 252:183–190. [PubMed: 20531000]

41. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT. Explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c869. [PubMed: 20332511]

TABLE I

Characteristics of Administrative Datasets and Registries Used for the Study of Pediatric Cancer

	Nationwide Inpatient Sample (NIS)	Kids Inpatient Database (KID)	Pediatric Health Information System (PHIS)	Surveillance, Epidemiology, and End Results Database (SEER)	National Cancer Data Base (NCDB)
Description	Publically-available nationally-representative all-payer inpatient care database	Publically-available nationally-representative all-payer inpatient care database for children	Privately administered database from >40 children's hospitals	Publically-available representative population-based registry of patients with cancer	Private cancer registry of patients at hospitals in Commission on Cancer
Data Source	State public and private data organizations	State public and private data organizations	Hospital encounters, including admissions, ambulatory medical and/or short-stay, ED visits	Public and private data organizations, cancer and hospital registries	Hospital cancer registries
Sample Data	Patient demographics	Patient demographics	Patient demographics	Patient demographics	Patient demographics
	Primary and secondary diagnoses, procedures	Primary and secondary diagnoses, procedures	Primary and secondary diagnoses, procedures	Disease incidence	Tumor characteristics
	Hospital characteristics	Discharge status	Date-stamped billing data	Tumor characteristics	Hospital characteristics
	Expected payment source	Hospital characteristics	Payer information	First-course treatment	Payer information
	Total charges	Expected payment source		Survival and vital status	First-course treatment
	Discharge status	Total charges			Survival
	Length of stay	Length of stay			
	Severity and comorbidity measures	Severity and comorbidity measures			
Limitations	Lacks clinical data, cannot track patients across time or settings	Lacks clinical data, limited tracking across time or settings	Coding errors, lack of clinical detail, limited classification of complications, limited tracking across time or settings	Some populations overrepresented, frequency of tumor characteristics and survival limited to the racial/ethnic groups for which population denominators available	Not population-based, excludes hospitals not in the Commission on Cancer