Clinical Versus Administrative Data

14

Sara K. Pasquali and J. William Gaynor

Abstract

Clinical and administrative databases have become increasingly utilized in the fields of pediatric cardiology and pediatric cardiac surgery for research, health policy activities, quality improvement, and evaluation of hospital performance. This chapter reviews the attributes of both types of datasets and discusses their strengths and weaknesses with regard to case ascertainment, risk adjustment, and outcomes assessment.

Keywords

Database • Outcomes • Congenital heart disease

Background

Large multicenter databases and registries have become increasingly utilized in the fields of pediatric cardiology and pediatric cardiac surgery over the past two decades [1]. These datasets serve several functions. First, they have allowed several different types of clinical research analyses including outcomes and comparative effectiveness type studies [2–4]. Due to the relative rarity of congenital heart disease, studies from a single center often lack statistical power or may have limited generalizability due to variation in practice and outcomes across centers. The use of large multicenter datasets helps to overcome these limitations. Health policy type analyses have also been conducted, such as investigations of the impact of center surgical volume or other structure and process measures on patient outcomes [5, 6]. In addition, large multicenter databases have been used for quality improvement purposes [7, 8]. As discussed in the previous chapters, several professional societies and other organizations now collect data across many centers regarding specific groups of patients with congenital heart defects, or those undergoing certain procedures. These data are then used to provide feedback to programs regarding outcomes benchmarked to national averages and

S.K. Pasquali, MD, MHS (✉)
Department of Pediatrics, C.S. Mott Children's Hospital, University of Michigan Congenital Heart Center, 1540 E. Hospital Drive 11-715z, Ann Arbor, MI 48109, USA
e-mail: pasquali@med.umich.edu

J.W. Gaynor, MD
Department of Cardiac Surgery, The Children's Hospital of Philadelphia, 34th St. and Civic Ctr. Blvd., Suite 12NW19, Philadelphia, PA 19104, USA
e-mail: gaynor@email.chop.edu

P.R. Barach, J.P. Jacobs, S.E. Lipshultz, P.C. Laussen (eds.), *Pediatric and Congenital Cardiac Care: Volume 1: Outcomes Analysis*,
DOI 10.1007/978-1-4471-6587-3_14,

peer institutions. This information may also be used for collaborative learning purposes where centers with the optimal outcomes are identified, best practices elucidated, and shared with other participating centers with the aim of improving overall quality of care and outcomes. Finally, information from large multicenter pediatric cardiac datasets has also been used for the purposes of public reporting of outcomes, ranking of hospital performance, and in some cases by large payers as evidence with which to base selective contracting, etc. [9–11].

Types of Datasets

In general there are two main types of datasets used for the purposes described above (Table 14.1).

Clinical Registries/Databases

These datasets are often run by professional societies or research groups for the purposes of quality improvement and/or research investigation. They are most often specific to a particular type of congenital heart disease, or collect information on patients undergoing specific types of procedures, e.g. congenital heart surgery, or cardiac catheterization. As described in the previous chapters, many of these types of datasets now exist in the field; examples include the Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database, and the American College of Cardiology Improving Pediatric and Adult Congenital Treatment (IMPACT) Registry [7, 12]. Clinical registries and databases collect detailed information using standardized definitions specific to patients with congenital heart disease. Outcomes data captured are most often focused on clinical outcomes and may not include resource utilization data. Data are usually collected and entered by trained data managers under the direction of the clinical care team, or in some cases directly by clinicians. While participation is usually "voluntary", there are generally guidelines regarding the inclusion/exclusion of cases, and many datasets perform audits to evaluate the inclusion of eligible cases, and assess the degree of accuracy of important variables [7].

Administrative Datasets

Administrative datasets contain information already being collected for the purposes of hospital billing or insurance claims. They are not specific to congenital heart disease and most often collect information regarding all hospitalized patients at a state or national level, or all patients covered by a certain payer or insurer. Alternatively, some employ sampling strategies that allow a fixed sample of patients or hospitals to represent the overall sample [13]. Examples of these types of datasets include the Children's Hospital Association Pediatric Health Information Systems (PHIS) Database, the

Table 14.1 Administrative vs. clinical data

	Type of dataset	
	Clinical	Administrative
Population	Specific to CHD patients	All hospitalized patients
Purpose	Run by professional societies or research groups for purposes of research or QI	Data collected for hospital billing purposes
Coding	CHD-specific codes	ICD-9 codes
Data collection	Trained data managers/clinicians	Billing personnel
Data		
CHD comorbidities/clinical outcomes	Detailed	Limited
Resource utilization data	Limited	Detailed

CHD congenital heart disease, *ICD* International Classification of Diseases, 9th Revision, *QI* quality improvement

Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project datasets such as the Kid's Inpatient Database (KID), and state Medicaid datasets. Administrative datasets most often capture information in the form of a Uniform Hospital Discharge Dataset, which is a uniform minimum dataset that captures information about a hospitalization including demographics, International Classification of Diseases, version 9 (ICD-9) diagnosis and procedure codes, outcomes such as mortality, and important resource utilization data such as hospital charges [1]. Some datasets such as the PHIS database capture additional resource utilization data such as utilization of medications, laboratory tests, imaging, etc. The codes and data captured in administrative data sources are not specific to congenital heart disease. The data are captured by hospital coding and billing professionals. Data capture and submission is generally mandatory, and inclusive by the nature of these types of datasets.

Strengths and Weaknesses of Clinical vs. Administrative Datasets

The relative merits of these different types of datasets have been debated for several years. In cardiac surgery, many concerns were initially raised in the 1980s when the U.S. Health Care Financing Administration (HCFA) sent reports to each US cardiac surgeon informing them of their outcomes based on analyses of administrative data [14]. While the validity of the data was questioned in many cases, there were no alternative data sources at the time that could be analyzed. This prompted the formation of the Northern New England Cardiovascular Disease Study Group, a group of cardiac surgeons and epidemiologists representing several programs in Northern New England. This group conducted pioneering work in the field to develop one of the first clinical registries designed to collect uniform information on cardiac surgery patients. In particular, given the concerns regarding the HCFA data, the group was interested in ensuring that all relevant cases were captured, and that data were captured to allow for accurate adjustment for important differences in patient characteristics and case mix when evaluating outcomes across institutions. The registry subsequently served as the foundation for collaborative learning and quality improvement, which led to an overall improvement in outcomes across the region [14]. Since that time, many other clinical registries have been developed by other groups, and the strengths and weaknesses of administrative and clinical data continue to be debated. In assessing the relative merits of these different types of datasets, there are several important issues to consider.

Case Ascertainment

The first important area to consider surrounds issues related to case ascertainment; this includes both whether the dataset captures all relevant cases (for example all patients undergoing congenital heart operations at a hospital), and also whether the cases are coded correctly. On the one hand, proponents of administrative data note that these datasets may be more generalizable since by design they are inclusive of all patients and programs either nationally or in a specific region or state [1]. In addition, it is argued that there may be less potential for "gaming the system" or omission of cases with less than optimal outcomes from inclusion in the dataset, as the individuals collecting and submitting administrative data are not being judged or evaluated based on the data, while this may not be the case with clinical registry data where practitioners whose performance is being evaluated may be involved in the collection and submission of the data. However, over the past several years as many clinical registries have expanded, issues related to generalizability have become less of a concern. For example, the STS Congenital Heart Surgery Database now represents more than 85 % of all US pediatric heart surgery programs [7]. Submission of all eligible cases is most often a stipulation of participation in a clinical registry, however mechanisms to ensure this are still under development in many cases. While some registries have audit

programs that aim in part to evaluate appropriate inclusion of all cases, this is not uniform across all registries.

The other important issue related to case ascertainment involves accurate coding of cases. It has long been known that ICD-9 codes do not cover the breadth and depth of congenital heart disease or procedures. For example, there is no ICD-9 code for the Norwood operation. Thus, a combination of various diagnosis and procedure codes are often used by investigators in order to attempt to identify patients undergoing certain types of operations, including the Norwood operation, in analyses using administrative data. In addition to problems with the codes themselves, while the administrative coding personnel are skilled at coding, they likely have limited knowledge of congenital heart disease, and do not have regular contact with the medical team to clarify conflicting data or inconsistent documentation in the medical record. Several groups have investigated the scope of this problem related to miscoding of cases. Our group recently performed an analysis of more than 55,000 children across 33 hospitals undergoing congenital heart surgery who each had information collected and coded in both a clinical registry (STS Congenital Heart Surgery Database), and an administrative dataset (PHIS Database) [15]. We compared the operation coded for each patient between datasets. Using the clinical registry data as the gold standard, we found that for four of eight benchmark operations analyzed, there was a greater than 10 % difference in the number of cases identified in the administrative vs. the clinical data. While the negative predictive value of the administrative data was high across operations (98.8–99.9 %), the positive predictive value was lower (56.7–88.0 %). This indicates that it is highly likely that a patient without a certain operation coded in the administrative data truly did not have that operation performed as assessed in the clinical registry. Conversely, the lower positive predictive value indicates that many patients coded as having undergone a certain type of operation in the administrative data are false positives and actually had a different operation performed based on evaluation of the clinical registry data. In addition to individual operations, we also evaluated categories of operations with similar mortality risk [the Risk Adjustment in Congenital Heart Surgery (RACHS) categories]. Overall, the percent agreement between the administrative and clinical registry data regarding RACHS category assignment was 68.4 %. We also found that there were relatively consistent findings across hospitals with regard to misclassification/miscoding, suggesting that these discrepancies are likely more related to the limitations of the ICD-9 system and coding methodology itself, rather than related to any hospital-specific factors such as volume or case mix [15]. Several other studies have also reported similar findings regarding differences in coding and case ascertainment between administrative and clinical datasets [16–18]. These findings may have implications for the use of administrative datasets in evaluation of the number and type of congenital heart operations or cases across institutions.

Risk Adjustment

A second important area to consider in evaluating the relative merits of clinical vs. administrative data involves risk adjustment. Accurate adjustment for potential differences in important patient characteristics and case mix is important in a variety of situations, including when comparing or reporting outcomes across hospitals, or when comparing groups of patients in an observational analysis. Because administrative datasets and the ICD-9 coding system are focused on the larger general population of hospitalized patients, they do not necessarily collect information regarding important comorbidities or patient characteristics that are more specific to patients with congenital heart disease. For example, it is well known that weight at surgery is an important factor impacting outcome following congenital heart surgery, particularly in neonates [19]. However this variable is not collected in many administrative datasets. While clinical registries often collect more specific data related to comorbidities and characteristics of patients with congenital heart disease, there are still certain limitations in that not every

potential variable of interest may be captured, particularly those that may only pertain to only a subset of diagnoses or procedures.

Another issue is related to "date-stamping." This refers to the fact that ICD-9 diagnosis codes do not differentiate between conditions present at admission vs. those that developed during the hospitalization and may be complications. Thus, there can be misidentification of post-operative complications as comorbidities, and vise versa. Some datasets have begun to address this by including a variable that indicates "present on admission."

Outcomes Assessment

A final topic of consideration relates to outcomes assessment. First, because of the general nature of administrative datasets and ICD-9 codes, they may not capture all outcomes of specific interest to the congenital heart disease population, for example certain post-operative complications. Second, studies in the adult literature suggest that there can be significant miscoding or misclassification of certain outcomes such as post-operative complications in administrative data. For example, a recent analysis compared records of patients with information collected both in a Medicare claims data set (an administrative dataset) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) registry [20]. Across 117,752 patients from more than 200 hospitals, investigators found that the sensitivity of the administrative data for detecting post-surgical complications coded in ACS-NSQIP ranged from 0.27 to 0.78 across various major complications. Although differences in complications have not been investigated extensively in the congenital heart surgery population, coding and capture of in-hospital mortality has been evaluated. In our recent analysis of more than 55,000 children across 33 hospitals who had data collected in both the STS Congenital Heart Surgery Database (clinical registry), and the PHIS Database (administrative dataset), we found that overall there was 99.83 % agreement between databases in in-hospital mortality, suggesting that there is not significant miscoding or capture of mortality data between these datasets [15].

In addition to the capture and correct coding of outcomes themselves, a second issue relates to errors in outcomes assessment due to miscoding or misclassification of cases. In other words, are there differences in the outcomes reported for certain diagnosis or procedure groups between datasets that are not related to the coding of the outcomes themselves, but related to differences in coding of the diagnosis or procedure across datasets? Our recent work with the STS and PHIS databases has investigated this further. We found that the differences in case ascertainment between data sources described above led to significant differences in outcomes assessment, for example an underestimation of mortality associated with truncus arteriosus repair by 25.7 % to an overestimation of mortality associated with ventricular septal defect repair by 31 % [15]. Differences were also found when evaluating mortality associated with larger groups of operations (the RACHS categories) between datasets, however these did not reach statistical significance. Importantly, only patients with concordant mortality status between the datasets were included in this analysis, in order to eliminate the possibility that any difference in outcomes identified might be related to differences in the coding of the outcomes themselves, rather than difference in the coding/classification of operations [15]. These findings may have implications for the use of administrative data in outcomes assessment, particularly at the level of individual congenital cardiac operations.

A final point to consider is the type of outcomes collected by clinical vs. administrative data sources. While clinical registries often collect more detailed clinical outcomes data, they most often do not collect the valuable resource utilization information contained in administrative datasets. In this era of increasing healthcare expenditures, it has become increasingly important to incorporate measures of cost into many analyses, as hospitals are under increasing pressure to not only optimize quality of care and outcomes, but also to reduce costs [21]. The

detailed resource utilization data contained in many administrative datasets is very valuable for these types of evaluations. Finally, neither type of dataset currently contains long-term follow up information, and most are focused primarily on in-hospital or short term outcomes. Methodology to incorporate longer-term clinical outcomes, resource utilization, and neurodevelopmental outcomes and quality of life is needed.

Conclusions

In summary, there are several important points to consider when evaluating the relative merits of administrative vs. clinical registry data, including issues related to case ascertainment, risk adjustment, and outcomes assessment. These factors may impact the relative utility of each type of dataset in outcomes and quality analyses, and in the evaluation of hospital performance. It is also important to note that while each type of dataset has its advantages and disadvantages, it is possible to capitalize on the strengths and mitigate some of the weaknesses of each dataset through database linkage strategies which allow for robust investigations not possible with either type of dataset alone [22]. This will be discussed further in Chap. 30.

References

1. Welke KF, Karmlou T, Diggs BS. Databases for assessing the outcomes of the treatment of patients with congenital and paediatric cardiac disease – a comparison of administrative and clinical data. Cardiol Young. 2008;18:137–44.
2. Pasquali SK, Hall M, Li JS, Peterson ED, Jaggers J, Lodge AJ, Marino BS, Goodman DM, Shah SS. Corticosteroids and outcome in children undergoing congenital heart surgery: analysis of the pediatric health information systems database. Circulation. 2010;122:2123–30.
3. Fudge JC, Li S, Jaggers J, O'Brien SM, Peterson ED, Jacobs JP, Welke KF, Jacobs ML, Li JS, Pasquali SK. Outcomes in patients with down syndrome undergoing congenital heart surgery: analysis of a national clinical database. Pediatrics. 2010;126:315–22.
4. Karamlou T, Hirsch J, Welke K, Ohye RG, Bove EL, Devaney EJ, Gajarski RJ. A united network for organ sharing analysis of heart transplantation in adults with congenital heart disease: outcomes and factors associated with mortality and retransplantation. J Thorac Cardiovasc Surg. 2010;140:161–8.
5. Pasquali SK, Jacobs JP, He X, Hornik CP, Jaquiss RDB, Jacobs ML, O'Brien SM, Peterson ED, Li JS. The complex relationship between center volume and outcome in patients undergoing the Norwood operation. Ann Thorac Surg. 2012;93:1556–62.
6. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. The relationship between hospital surgical case volumes and mortality rates in pediatric cardiac surgery: an national sample, 1988–2005. Ann Thorac Surg. 2008;86:889–96.
7. Jacobs JP, Jacobs ML, Mavroudis C, Lacour-Gayet FG, Tchervenkov CI, Pasquali SK. Executive summary: the Society of Thoracic Surgeons Congenital Heart Surgery Database – 18th Harvest – (January 1, 2009–December 31, 2013). The Society of Thoracic Surgeons (STS) and Duke Clinical Research Institute (DCRI), Duke University Medical Center, Durham, Spring 2013 Harvest.
8. Kugler JD, Beekman RH, Rosenthal GL, Jenkins KJ, Klitzner TS, Martin GR, Neish RS, Lannon C. Development of a pediatric cardiology quality improvement collaborative: From inception to implementation. From the Joint Council on Congenital Heart Disease Quality Improvement Task Force. Congenit Heart Dis. 2009;4:318–28.
9. Pediatric Congenital Cardiac Surgery in New York State 2006–2009. Available at: http://www.health.ny.gov/statistics/diseases/cardiovascular/heart_disease/docs/2006-2009_pediatric_congenital_cardiac_surgery.pdf. Accessed 10 Jul 2013.
10. U.S. News & World Report's Annual Ranking of Best Hospitals. Available at: http://www.rti.org/besthospitals. Accessed 28 May 2013.
11. OptumHealth Congenital Heart Disease Center of Excellence. Available at: https://www.myoptumhealthcomplexmedical.com/gateway/public/chd/providers.jsp. Accessed 28 May 2013.
12. Martin GR, Beekman RH, Ing FF, Jenkins KJ, McKay CR, Moore JW, Ringel RE, Rome JJ, Ruiz CE, Vincent RN. The IMPACT registry: improving pediatric and adult congenital treatments. Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu. 2010;13:20–5.
13. The Nationwide Inpatient Sample Database. Available at: http://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp. Accessed 10 Jul 2013.
14. Likosky DS. Lessons learned from the Northern New England Cardiovascular Disease Study Group. Prog Pediatr Cardiol. 2012;33:53–6.
15. Pasquali SK, Peterson ED, Jacobs JP, He X, Li JS, Jacobs ML, Gaynor JW, Hirsch JC, Shah SS, Mayer JE. Differential case ascertainment in clinical registry vs. administrative data and impact on outcomes assessment in pediatric heart surgery. Ann Thorac Surg. 2013;95:197–203.
16. Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, et al. The importance of nomenclature for congenital cardiac disease: implications for research and evaluation. In: Jacobs JP, editor. 2008 Cardiology in the

Young Supplement: Databases and The Assessment of Complications associated with The Treatment of Patients with Congenital Cardiac Disease. Prepared by: The Multi-Societal Database Committee for Pediatric and Congenital Heart Disease. Cardiology in the Young, vol. 18, issue S2 (Suppl 2), p. 92–100, December 9, 2008.
17. Cronk CE, Malloy ME, Pelech AN, et al. Completeness of state administrative databases for surveillance of congenital heart disease. Birth Defects Res A Clin Mol Teratol. 2003;67:597–603.
18. Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, Falken MC. Validity of hospital discharge data for identifying infants with cardiac defects. J Perinatol. 2005;25(11):737–42.
19. Curzon CL, Milford-Beland S, Li JS, O'Brien SM, Jacobs JP, Jacobs ML, Welke KF, Lodge AJ, Peterson ED, Jaggers J. Cardiac surgery in infants with low birth weight is associated with increased mortality: analysis of the Society of Thoracic Surgeons Congenital Heart Surgery Database. J Thorac Cardiovasc Surg. 2008;135:546–51.
20. Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L, Rapp M, Co CY. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. Ann Surg. 2012;256:973–81.
21. Pasquali SK, Gaies MG, Jacobs JP, Gaynor JW, Jacobs ML. Centre variation in cost and outcomes for congenital heart surgery. Cardiol Young. 2012;22:796–9.
22. 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;160:1099–104.